AD_____

Award Number: DAMD17-03-1-0657

TITLE: Multiple Aperture Radiation Therapy (MART) for Breast Cancer

PRINCIPAL INVESTIGATOR: Tianfang Li, Ph.D.

CONTRACTING ORGANIZATION: Stanford University
Stanford, CA 94305

REPORT DATE: November 2005

TYPE OF REPORT: Annual Summary

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
01-11-2005	Annual Summary	1 Nov 2004 – 31 Oct 2005

4. TITLE AND SUBTITLE

Multiple Aperture Radiation Therapy (MART) for Breast Cancer

5a. CONTRACT NUMBER

5b. GRANT NUMBER
DAMD17-03-1-0657

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Tianfang Li, Ph.D.

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Stanford University
Stanford, CA 94305

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES
Original contains colored plates: ALL DTIC reproductions will be in black and white.

14. ABSTRACT

Compared with conventional methods for breast cancer RT such as opposed tangential fields (OTF), modern IMRT provides unprecedented potential to deliver 3D-dose distributions with superior tumor conformality and normal tissue spare. In practice, however, the current IMRT process deviates significantly from the conventional approach and requires considerable costs of the time and health care. The complexity of the treatment and deliver procedure make this technology problematic for the breast radiotherapy. We have carried out a systematic study on a new solution, which is a hybrid of the advantages of the IMRT and OTF, termed as multiple aperture radiotherapy (MART). The efficiency in dose shaping and simplicity in dose delivery make it a perfect choice for breast cancer treatment.A number of important milestones have been accomplished, which include (i) developed an efficient optimization method for MART inverse planning; (ii) implemented MART in a clinic environment for breast treatment; (iii) proposed a gated IMRT/MART based on the cutting-edge technology of 4D CT imaging; (iv) established a robust technique for low-dose 4D CT acquisition to reduce the risks of radiation-induced cancer. It is expected these tools will greatly facilitate the imaging, planning, delivery, and quality assurance of breast radiation treatment.

15. SUBJECT TERMS
Breast Cancer

16. SECURITY CLASSIFICATION OF:				18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER *(include area code)*
U	U	U	UU	53	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

Table of Contents

INRTODCUCTION

This postdoctoral traineeship grant (DAMD17-03-1-0657, entitled "Multiple Aperture Radiation Therapy for Breast Cancer") was awarded for the period of Nov. 1, 2003 to Oct. 31, 2005, and was transferred to the current principal investigator (PI) on Nov. 1, 2004. The goal of this project is to develop a novel radiation treatment technique, multiple aperture radiation therapy (MART), as a candidate modality for treating breast cancer. Under the generous support from the U.S. Army Medical Research and Materiel Command (USAMRMC), the PI has gained a tremendous amount of knowledge on breast cancer and breast cancer management. The support has also made it possible for the PI to contribute significantly to breast cancer research. A number of conference and refereed publications have been resulted from the support. This annual report is the final report to conclude the scientific results and achievements made during the work of the project.

Conventional radiotherapy for breast cancer utilizes two opposed tangential fields (OTF) with either uniform or wedged photon beams [1-4], and often presents problems related to breast dose inhomogeneity and relatively high doses to the ipsilateral lung and heart, because certain volume of the ipsilateral lung and, in the case of the left breast, part of the heart is inevitably included in the tangential fields. As a consequence, breast irradiation has been associated with a number of potential complications [5-8], including radiation induced pneumonitis, cardiac toxicity, rib fracture, arm edema, severe breast or chest wall fibrosis, soft tissue or bone necrosis, and radiation induced secondary cancer. Intensity Modulated Radiation Therapy (IMRT) is a highly promising technology that can potentially overcome these problems [9-14]. However, there are several practical issues in the treatment planning, delivery and quality assurance (QA) processes, which add significant overhead as compared to the standard OTF and make it clinically less useful. For instance, current IMRT treatment planning requires physicians to segment explicitly the target volume, which takes in average about 20 minutes per patient and increases the cost of health care. The procedure of dosimetric verification of an complicate shape intensity-modulated beam is yet not well established and needs great effort of quality control. In addition, the IMRT optimization algorithm is not dedicated to breast cancer and may not be the most efficient scheme. In short, an effective method to plan and deliver IMRT breast treatment is highly desired in order for tens of thousands of breast cancer patients to benefit from the state-of-the-art technology. For this purpose, we have proposed the MART approach in this USAMRMC-supported project.

The two most important aspects in radiation treatment technique such as IMRT are the conformality of the radiation to the target and the uniformity of the radiation inside the target volume. To achieve these goals while maintaining the simplicity of conventional radiation technique such as the OTF, MART has been designed by adding finite beam segments to a properly selected tangential field so that the dose inhomogeneity resulted from OTF is efficiently removed and the total dose are better confined to the target volume. The specific aims of this work are: (a) to demonstrate that MART can lead to substantially improved dose distribution over the conventional method, without increasing the complexity of the radiation treatment; (b) to show that the superior dose distributions of MART can be realized efficiently and accurately. Since the respiration of patient can affect the accuracy of the radiation treatment for breast cancers, techniques related to the motion compensation or four-dimensional (4D) treatment are extensively studied. The details are given below.

BODY

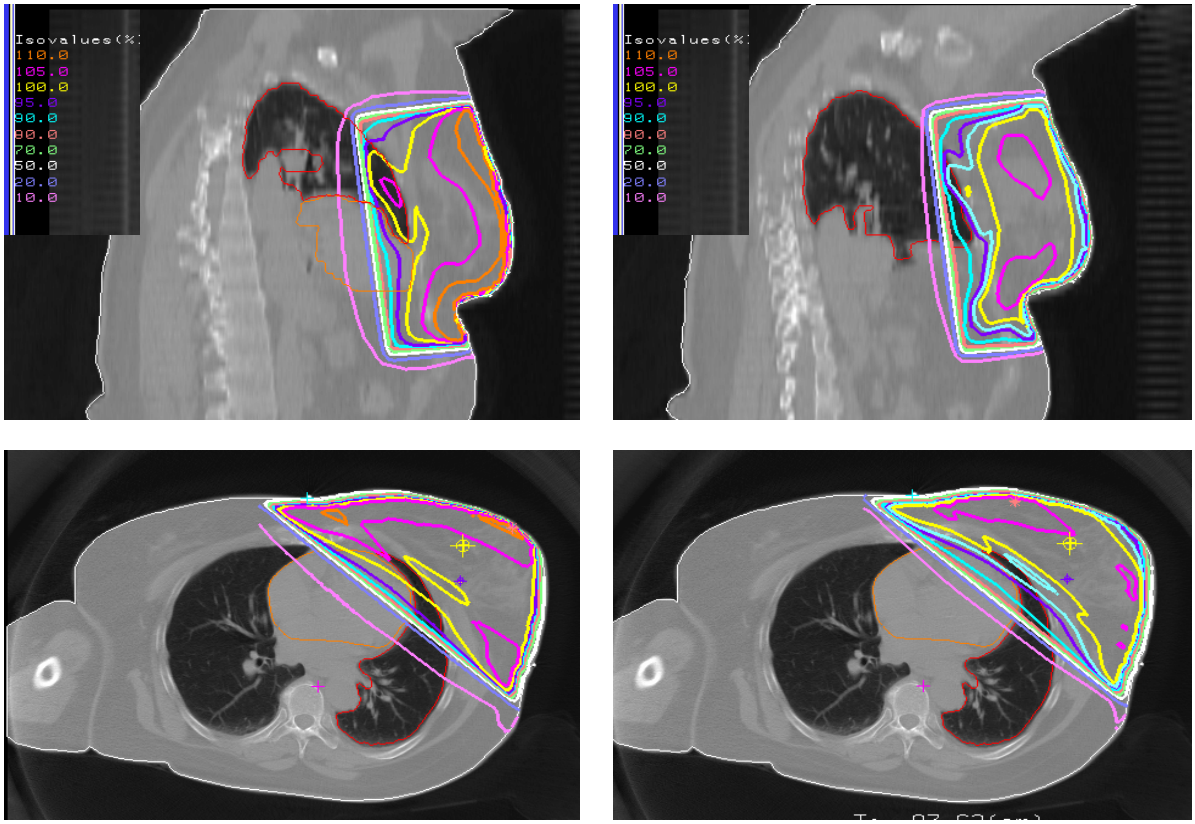
A. MART Planning

We have used a commercial inverse planning system (Corvus, North American Scientific, Cranberry Township, PA) to generate MART plans for a number of breast treatments at Stanford University Hospital. Several closely related clinical issues, including the type of wedges (physical or dynamic), field size, incorporation of MLC transmission into the step-and-shoot delivery, and QA, were extensively investigated during the implementation of the MART technique. Our study indicated that the MART markedly improves breast irradiation and provides superior dose distributions needed to reduce radiation side effects and complications. The technique is especially valuable for radiation treatment of large-breasted women, where it is difficult to achieve homogeneous target dose distribution.

The MART patients underwent computed tomography (CT) in the conventional treatment position supported by an Alpha Cradle immobilization device (Smithers Medical Products, Tallmadge, Ohio). Radiopaque markers were placed on the patients' chest to indicate the medial and lateral borders of the palpable breast tissue and the location of the lumpectomy scar. The radiation-sensitive structures included the left and right lungs, the heart, and the contralateral breast. For the purpose of quantitative study, the contours of the skin, target volume and the sensitive structures were outlined using the segmentation tools provided by the virtual simulation workstation (AcQSim^{TM}, Philips Medical System, Cleveland, OH). It is, however, not required to outline the structures in general MART treatment. The tangential fields were determined by the routine virtual simulation procedure performed on an AcQSim workstation. The fields may be adjusted at the stage of treatment planning according to the actual treatment objective for each patient with considerations concerning tumor bed coverage, in-field lung and cardiac volume, if it is left-breast irradiation. The MART planning was done with a 3D treatment planning system (FOCUS^{TM}, Computerized Medical System, St. Louis, MO). The MART planning started with a standard OTF plan, then additional MLC field segments were introduced to one or both beam directions to boost the "cold" region(s) under the guidance of dose distributions in the plane perpendicular to the incident beam direction. The weights and MLC apertures of the segments were adjusted manually through trial-and-error process to achieve a uniform dose distribution. Our experience indicated that, for intermediately complex cases, it was often sufficient to introduce one or two additional segments to the original opposed tangential fields. For complex cases, two or three additional segments were frequently used. A segment with ipsilateral lung or heart blocked by MLC (the third segment in either medial or lateral MART field) was also helpful in reducing the dose to these structures.

Fig. 1 showed a typical clinical case that fell into the category of intermediately complicated or complicated cases. For comparison, two plans were generated. One was the standard OTF plan and the other one was the MART plan. The prescribed dose was specified to a point ~3 cm anterior to the isocenter and it was desired that the 100% isodose curve to cover the breast target volume. The prescription dose of all treatment plans were scaled to 5,040 cGy. The hot spots of the standard OTF plans in the breast volume ranged from 109% to 118% when normalized to the prescribed dose. As shown in Fig. 1, the isodose distributions in the central transverse section and in a plane perpendicular to the incident beams for the standard OTF and MART treatments, indicate that the dose inhomogeneity in the target volume was significantly reduced with MART, as well as reduction in the high dose to the ipsilateral lung and heart when compared with the OTF plan. The target maximum dose was reduced from 118% to 112% for the MART plan. Furthermore, the target volume receiving high dose irradiation was significantly

reduced. In order to include the medial breast tissue into the radiation field, ~10% of the heart volume and the left lung were included in the tangential fields. As thus, a significant fraction of the heart and lung receives high radiation dose in standard OTF plan. The high doses to the heart was reduced by almost 6% using MART technique as a result of adding one additional segment in each incident beam, together with ~5% improvement in the maximum target dose in the target volume. The heart volume and ipsilateral lung volumes receiving high dose irradiation were also markedly reduced for MART treatment.



(a) (b)

Figure 1. Comparison of the isodose distributions of the treatment plans of the left-sided breast case using the tangential field technique (a), MART (b). Target volume includes the whole breast and the internal mammary nodes. Isodose levels are shown at 110%, 100%, 90%, 70%, 50%, 30%, and 10%.

B. Optimization with Voxel Dependent Priori and Genetic Algorithm for IMRT/MART

In developing an effective aperture based optimization algorithm for breast irradiation, it is found that ALL current inverse planning algorithms treat all voxels within a target or sensitive structure equally and use structure specific prescriptions and weighting factors as system parameters. In reality, the voxels within a structure are not identical in complying with their dosiemtric goals and there exists strong intra-structural competition in these voxels. Inverse planning objective function should not only balance the competing objectives of different structures but also that of the individual voxels within various structures.

We have proposed an approach to improve IMRT/MART treatment plan by purposely modulating the penalty on individual voxel level based on the *a priori* dosimetric capability information of the dose optimization system. We quantify the degree for a voxel to achieve its dosimetric goal by introducing the concept of dosimetric capability for each voxel in a target or

sensitive structure. A genetic algorithm (GA) for breast MART inverse planning has been developed for optimizing simultaneously the shapes of a pre-specified number of segments and their corresponding weights. The GA encodes potential solutions in chromosome-like structures and applies different recombination operators (crossover and mutation) to explore the search space. The system variables, which include the weights of the segments and the leaf positions defining the shapes of the segments are encoded in three chromosome-like structures for each potential solution: the 1st one for the weights of the segments, the 2nd one for the positions of left bank of MLC leaves, and 3rd one for the right bank. All encodings are realized in integer format. The quality of a solution is evaluated according to its fitness, which is defined as the inverse of the usual quadratic objective value. The chance of a particular solution to be part of the next generation population is proportional to its survival probability, defined as the ratio of its fitness and the total fitness of the population. A more fit solution has a higher probability of being selected into the next generation. The new population is then selected by simulating the spinning of a suitable roulette wheel, *N* times, where *N* equals the number of solutions in the current generation. The selection process is followed by crossover and mutation operations, where the potential solutions interchange information that usually leads to improved individuals. In addition to the fitness-based selection process, we allow the best member of the current population to be automatically copied into the next generation. This process called elitism leads to a faster convergence and keeps track of the best solutions obtained in each iteration.

The segment-based optimized plans improved significantly the target dose uniformity in comparison with the standard OTF plans. The overall planning and treatment delivery overhead of the approach is significantly reduced compared to the conventional beamlet-based IMRT breast irradiation. For all cases (5 left and 5 right breast cancer patients) we have tested, it was found that 3~10 segments per tangential beam were sufficient to ensure highly homogeneous doses within the target. Our results also revealed that the maximum target dose could easily be reduced from 109%~117% in conventional OTF to 106% to 112%. The volume receiving high dose irradiation in the breast target was also markedly reduced. It was also possible to use segment-based optimization to reduce the dose to the ipsilateral lung/heart.

C. Motion Correction and Advanced 4D Imaging

In radiation therapy, respiratory motion poses significant challenges for tumors in breast. The motion can distort the shape of an object, degrade the anatomic position reproducibility during imaging, and necessitate larger margins during radiotherapy planning. It also causes inaccuracy in estimating the tumor volume, thereby preventing an effective dose escalation for the treatment of a target tumor. These issues make it difficult to achieve the desired goals of conformal radiotherapy. Four-dimensional (4D) CT scans, acquired synchronously with a respiratory signal, provide not only the three-dimensional (3D) spatial information, but also temporal changes of the anatomy as a function of the respiratory phase during the imaging, and can therefore be employed in 4D treatment planning to explicitly account for the respiratory motion, for example, the respiration gated IMRT/MART treatment, where the photon beams are only switched on in particular phase window determined by a real-time motion tracking system (RPM).

We have recently implemented a gated irradiation technique for the treatment of left-sided breast cancer, where the goal is to better conform the tumor target while sparing the heart. In figure 2 we show an example of the treatment. Fig. 2a and 2b show the inhale and exhale phases, respectively. Fig. 2c shows the plan for gated treatment of the patient. As seen from the isodose plot, the heart of the patient is spared greatly as compared with conventional free-

breathing treatment. Figure 2d shows the patient's breathing pattern. A temporal margin of 20% (the portion highlighted by thicker lines on the breathing curve) was imposed during the gated radiation delivery process.

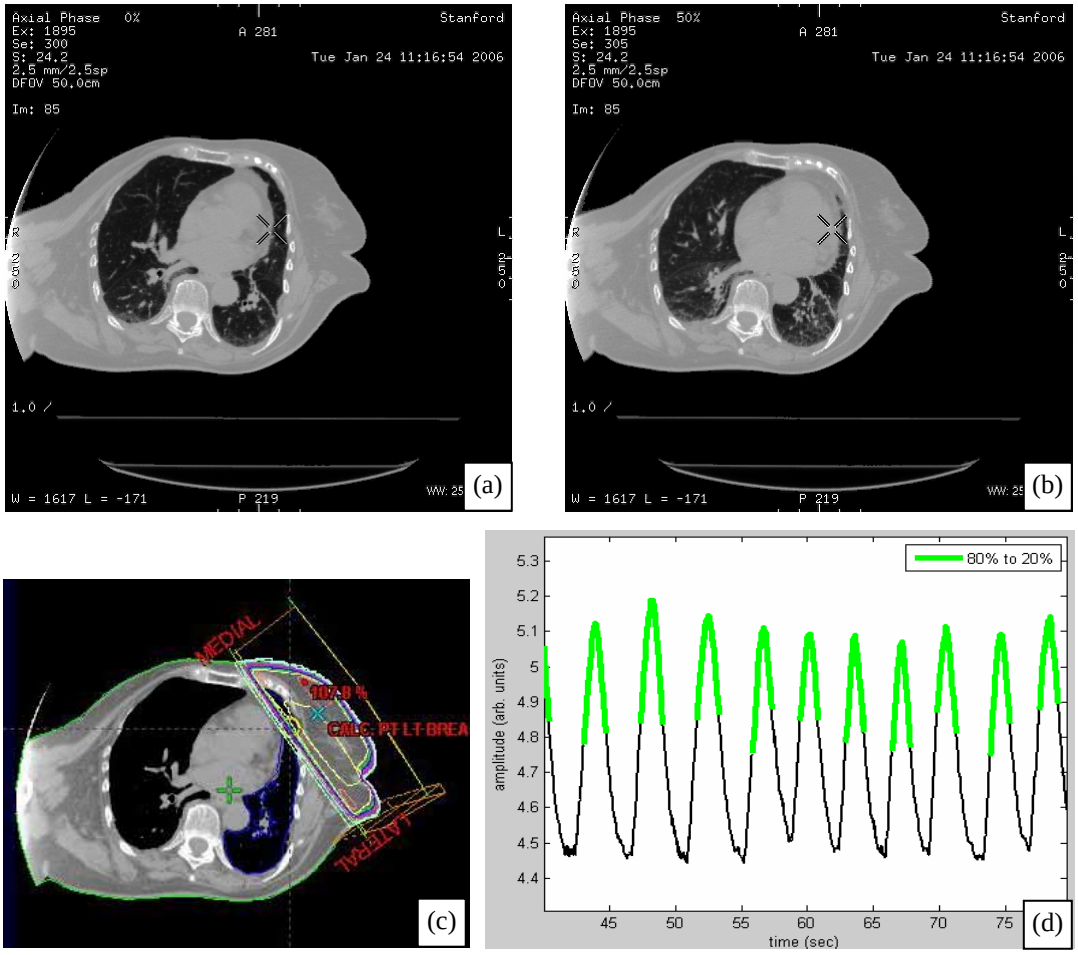


Figure 2. Gated breast radiation treatment with 4D CT. (a) and (b) are inspiration and expiration phases, respectively. (c) is the gated treatment plan, and (d) shows the patient breathing pattern, where the thick green line indicates the 20% to 80% respiratory phase window.

The above gated treatment requires a 4D simulation CT. An astonishing problem with 4D CT for breast cancer patients is the extremely high radiation dose during the 4D imaging (about 10 to 20 times higher than a regular 3D CT scan). In particular for patients with partial breast treatment, such high dose screening may lead to secondary radiation-induced cancer in the normal tissue (the other normal breast for example). One of our aims in this projection therefore became to lower the radiation exposure to the patient in 4D imaging while maintaining the advantage of the provided kinetic information of the tumor and organs at risk (OAR) by 4D CT.

The method we developed in this project is to perform 4D CT scans at relatively low current, hence reducing the radiation exposure of the patients. To deal with the increased statistical noise caused by the low current, a novel 4D penalized weighted least square (4D-PWLS) smoothing method has been proposed, which can incorporate both spatial and phase information. The 4D images at different phases were registered to the same phase via a *deformable* model, thereby, a regularization term combining temporal and spatial neighbors can be designed for the 4D-PWLS objective function. The proposed method was validated with phantom experiments and applied to patient studies, where superior noise suppression and resolution preservation were observed.

4D CT images were acquired with a combined PET/CT scanner (Discovery ST/LightSpeed 8-slice, General Electric Medical Systems) in our clinic. Fig. 3 show the phantoms used for the validation of our 4D-PWLS method: one is a commercial calibration phantom CatPhan® 600 (The Phantom Laboratory, Inc., Salem, NY), and the other is an anthropomorphic thorax phantom. To acquire the 4D CT data, each phantom was placed on the top of a platform capable of sinusoidal motion along the cranial–caudal direction. A Real-time Position Management (RPM) respiratory gating system (Varian Medical Systems, Palo Alto, CA) was used to record the motion by tracking two infrared reflective markers, rigidly mounted on a plastic block on the top of the phantom, by means of an infrared video camera mounted on the PET/CT table. The clinical 4D-CT patient studies were performed on the same scanner. Patients were asked to breathe normally during the scan. The plastic block with two infrared reflective markers was taped on the top of the patient's abdomen, placed medially and a few cm inferior to the xiphoid processes. The respiratory signal of the patient from the RPM was recorded and synchronized with CT data acquisition, similar to the phantom scan. After the scan data were prospectively reconstructed at the PET/CT scanner, both the CT images and the corresponding motion data recorded by the RPM system were transferred to a GE Advantage Workstation (GE Medical Systems, Waukesha, WI). The "Advantage 4D" software on the workstation simultaneously displays the CT images and the motion data, and sorts the cine images into a set of respiratory phase images.

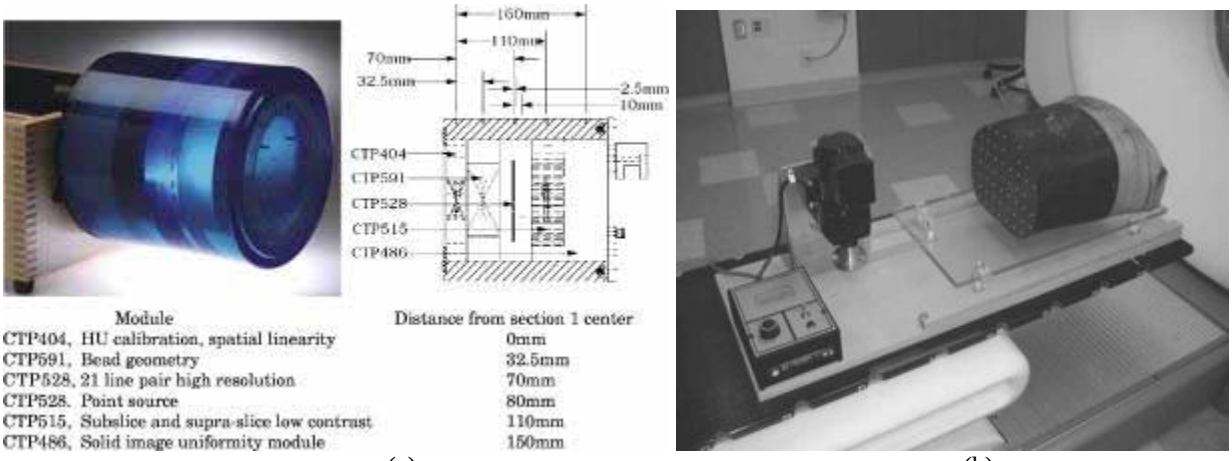


Figure 3. 4D CT phantoms (a) CatPhan, (b) anthropomorphic thorax phantom and the motion platform.

In Fig. 4 we compared the images for one slice of the CatPhan® 600 that contains multiple strips on a circle, of which the resolution sections are ranging from 1 to 21 lines pairs per cm. The left panel in the figure is the conventional 4D CT images acquired at 10 mA, the middle is the image obtained with the new method, and the right is the vertical profiles across the strips highlighted by a rectangle. The study showed that the new method improved the signal-to-noise (SNR) about 2.5x, with only a 6% loss in spatial resolution.

In the human thorax phantom study, we compared the smoothed CT images with the original low-mA (10 mA) CT images, as well as the original high-mA (100 mA) CT images. Fig. 5 shows their coronal images. The left column contains the 4D CT image acquired at 100 mA after sorting by the GE 4D Advantage software. The middle column shows the image at the same phases acquired at 10 mA, and image from our 4D-PWLS method based on the low-current CT data are displayed in the right column. The effective reduction of the image noise can

be observed, where the average SNR of 10 mA images increased by more than three-fold from 0.051 to 0.165 after the proposed post-processing.

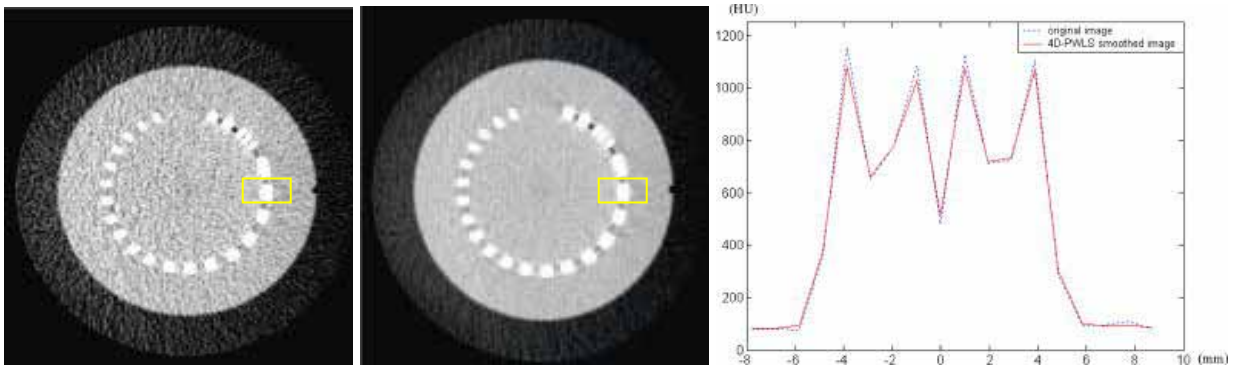


Figure 4. 4D-PWLS study with CatPhan® 600 phantom. The CTP528 section (high-resolution line pairs) is compared for 10 mA, where the left and middle are the images before and after the proposed 4D-PWLS processing, respectively. The profiles show the resolution change after the 4D-PWLS processing over the line pairs.

Figure 5. Phantom study for the 4D-PWLS method with the thorax phantom. The left and middle columns are the original phases obtained from the GE Advantage Workstation, for 100 and 10 mA, respectively; the right column shows the 10 mA phases after 4D-PWLS processing. The red rectangles represent the selected ROI for calculation of SNRs, containing 5×5×5 voxels.

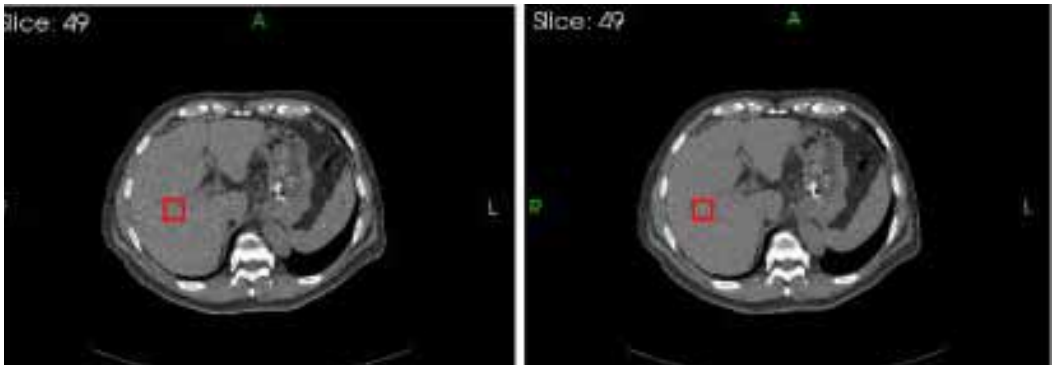


Figure 6. One example of the patient studies for the 4D-PWLS method. The phase shown here is the end-inspiration phase. The left is the regular image obtained from the GE 4D Advantage Workstation, and the right shows the image after 4D-PWLS processing. The red rectangles represent the selected ROI for the calculation of SNRs, each of which contain 5×5×5 voxels.

One example of patient studies is shown in Fig 6. The left column is the original phase image (end-inspiration phase) obtained from GE Advantage Workstation, and the right column is the image after processing with the proposed 4D-PWLS enhancement. Successful noise suppression is observed. For this patient, the SNRs increased from 2.204 to 4.558 for the end-expiration phase, and from 1.741 to 3.862 for the end-inspiration phase in the selected ROIs.

KEY RESEARCH ACCOMPLISHMENTS

First Year:

- Developed a manual MART planning procedure and assessed the dosimetric improvement resulted from MART, and demonstrated the advantage of MART over conventional OTF plans.
- Clinically implemented MART treatment technique and over hundred breast cancer patients have been treated using the new technique.

Second Year:

- Improved inverse planning framework by introducing a voxel dependent penalty and developed a dedicated genetic algorithm for MART planning. Demonstrated its superiority over conventional treatment planning systems in terms of dose homogeneity and efficiency.

- Developed a novel strategy of gated IMRT/MART treatment for breast caner. Constructed a breathing motion phantom for the 4D imaging/treatment research.

- Developed a method that significantly reduced the radiation exposure to the patient during the 4D CT imaging, which is an essential step in 4D radiation therapy for breast cancer.

REPORTABLE OUTCOMES

The following is a list of publications resulted from the grant support. Copies of the publication materials are enclosed with this report.

Refereed Publication:

- **Li T**, Schreibmann, E, Yang Y and Xing L, "Motion correction for improved target localization with on-board cone-beam computed tomography," *Physics in Medicine Biology* 51, 253-267, 2006.
- **Li T**, Schreibmann E, Thorndyke B, Tillman G, Boyer A, Koong A, Goodman K and Xing L, "Radiation dose reduction in 4D computed tomography," *Medical Physics* 32(12), 3650-3660, 2005.
- **Shou Z,** Yang Y, Cotrutz **C** and Xing L, "Quantitation of the *a priori* dosimetric capabilities of spatial points in inverse planning and its significant implication in defining IMRT solution space," *Physics in Medicine and Biology* 50(7), 1469-1482, 2005.

Published Abstracts:

The PIs' group has also been active in disseminating our research results. The following are some of the presentations given in various national/international meetings.

- **Li T,** Cotrutz C, Gofinett D, and Xing L, "Segmenation-based breast IMRT using a genetic dose optimization algorithm," *Era of Hope (Department of Defense Breast Cancer Research Program Meeting)* 131, 2005.

- Yang Y, **Li T**, Schreibmann, Boyer A, and Xing L, "Is cone beam CT suitable for dose verification?" Medical Physics 32, 2160, 2005.
- **Li T**, Yang Y, Schreibmann, and Xing L, "A new cone-beam CT repositioning technique through deformable registration," Medical Physics 32, 2160, 2005.
- Xing L, Schreibmann, Yang Y, Boyer A, and **Li T**, "Image segmenatation in 4D CT based on a deformable image registration model," Medical Physics 32, 2095, 2005.
- Kim G, **Li T**, Yang Y, Schreibmann, Thorndyke B, Boyer A, and Xing L, "Influence of respiratory motion on cone-beam CT imaging of thorax and abdomen," Medical Physics 32, 1934, 2005.
- Thorndyke B, Schreibmann, **Li T**, Boyer A, and Xing L, "A comparison of amplitude- and phase-based 4D CT," Medical Physics 32, 1919, 2005.
- Lo A, **Shou Z** and Xing L, Quantitative comparison of aperture-based and beamlet-based inverse planning techniques, *International Journal of Radiation Oncology, Biology, Physics, Volume 60, 1, Supplement 1, September 2004, Pages S630-S630;*
- **Shou Z** and Xing L, Aperture-based IMRT inverse planning with incorporation of organ motion**,** *International Journal of Radiation Oncology, Biology, Physics, Volume 60, 1, Supplement 1, September 2004, Pages S631-S632.*
- Xing L and **Shou Z,** Intrinsic spatial heterogeneity in inverse planning and its significant role in defining the universe of IMRT solution space, *International Journal of Radiation Oncology, Biology, Physics, Volume 60, 1, Supplement 1, September 2004, Pages S635-S635.*
- **Shou Z** and Xing L: Improve IMRT distribution by using spatially non-uniform important factors, oral presentation in 2004 annual meeting of ICCR in Seoul, Korea. 120-123, 2004.
- Xing L, Hunjan S, Lian J, Yang Y, **Shou Z**, Schreibmann E, Boyer A, Towards Biologically Conformal Radiation Therapy: Functional and Molecular Image Guided Intensity Modulated Radiation Therapy, XIVth International Conference on the Use of Computers in Radiation Therapy (ICCR), Soul, Korea, 2004.

CONCLUSIONS

Despite the well-appreciated fact that intensity-modulation could lead to significantly improved dose distributions in breast irradiation, its clinical implementation has been hindered by the deficiencies in the current inverse planning system and by the lack of a comprehensive treatment procedure. This is evidenced by that very few institutions are using IMRT routinely for breast cancer treatment. A challenge here is how to achieve intensity modulation without increasing the breast treatment complexity. In this work, a variant form of IMRT, MART, is developed and its utility for breast irradiation is evaluated. Delivery schemes with physical or dynamic wedge are addressed to meet the requirements of different clinical environments. It was found that, with the use of MART technique, the commonly seen high doses to the ipsilateral lung and, in the case of the left-breast cancer patient, the heart were reduced and the dose homogeneity in the breast target volume was significantly improved. The new treatment scheme is especially valuable for treating patients with large-sized breasts.

Segment-based IMRT for breast is a natural "breeding" of standard OTF and IMRT techniques that improve the dosimetry without paying excessive overhead associated with the current IMRT. Each segment is a standard field and conventional methods can be used for dose calculation and quality assurance. It is found that 3~10 segments per tangential beam are sufficient to ensure a highly homogeneous dose and good sparing of the ipsilateral lung/heart.

Furthermore, the approach eliminates the intermediate IMRT leaf sequencing process and leads to a directly deliverable solution. A novel voxel dependent penalty scheme was proposed to improve the currently existing dose optimization scheme. Specifically, the Genetic Algorithm fits well for the MART optimization tasks. The manual MART treatment have been implemented and now part of routine practice at Stanford University Hospital.

The cutting-edge technique of 4D treatment uses 4D imaging to achieve the kinetic information of the patient anatomy. It has great potential to breast cancer patient because of smaller margin and higher dose escalation may be used. The 4D imaging however requires 10 to 20 times higher radiation exposure for the scan than conventional way. Our work provided the clinicians an efficient tool to significantly reduce the radiation to patients while achieving the same quality of image. The new technique would be beneficiary to the 4D radiation therapy field as a whole.

REFERENCES:

1. Fisher B, Redmond C, Poisson R, et al.: Eight-year results of a randomized clinical trial comparing total mastectomy and lumpectomy with or without irradiation in the treatment of breast cancer. New England Journal of Medicine 1989; 320(13): 822-8.

2. Fisher B, Costantino J, Redmond C, et al.: Lumpectomy compared with lumpectomy and radiation therapy for the treatment of intraductal breast cancer. New England Journal of Medicine 1993; 328(22): 1581-6.

3. Dobbs HJ: Radiation therapy for breast cancer at the millennium. Radiotherapy & Oncology 2000; 54: 191-200.

4. Asrari F, Gage I: Radiation therapy in management of breast cancer. Current Opinion in Oncology 1999; 11(6): 463-7.

5. Arbetter KR, Prakash UB, Tazelaar HD, Douglas WW: Radiation-induced pneumonitis in the "nonirradiated" lung. Mayo Clinic Proceedings 1999; 74(1): 27-36.

6. Shapiro CL, Hardenbergh PH, Gelman R, et al.: Cardiac effects of adjuvant doxorubicin and radiation therapy in breast cancer patients. Journal of Clinical Oncology 1998; 16(11): 3493-501.

7. Rutqvist LE, Lax I, Fornander T, Johansson H: Cardiovascular mortality in a randomized trial of adjuvant radiation therapy versus surgery alone in primary breast cancer [see comments]. Comment in: Int J Radiat Oncol Biol Phys 1992;22(5):1157-8. International Journal of Radiation Oncology, Biology, Physics 1992; 22(5): 887-96.

8. Rothwell RI, Kelly SA, Joslin CA: Radiation pneumonitis in patients treated for breast cancer. Radiotherapy & Oncology 1985; 4(1): 9-14.

9. Karlsson M, Zackrisson B: Matching of electron and photon beams with a multi-leaf collimator. Radiotherapy & Oncology 1993; 29(3): 317-26.

10. Li JG, Williams SS, Goffinet DR, Boyer A, Xing L: Breast-conserving radiation therapy using combined electron and IMRT technique. Radiotherapy & Oncology 2000; 56: 65-71.

11. Smitt MC, Li SD, Shostak CA, Chang W, Boyer AL: Breast-conserving radiation therapy: potential of inverse planning with intensity modulation. Radiology 1997; 203(3): 871-6.

12. Hong L, Hunt M, Chui C, et al.: Intensity-modulated tangential beam irradiation of the intact breast. International Journal of Radiation Oncology, Biology, Physics 1999; 44(5): 1155-64.

13. Kestin LL, Sharpe MB, Franzier RC, et al.: Intensity-modulation to improve dose uniformity with tangential breast radiotherapy: initial clinical experience. International Journal of Radiation Oncology, Biology, Physics 2000; 48: 1559-1568.

14. Lo YC, Yasuda G, Fitzgerald TJ, Urie MM: Intensity modulation for breast treatment using static multi-leaf collimators. International Journal of Radiation Oncology, Biology, Physics 2000; 46(1): 187-94.

Appendix I. Copy of manuscripts

Radiation dose reduction in four-dimensional computed tomography

T. Li, E. Schreibmann, B. Thorndyke, G. Tillman, A. Boyer,
A. Koong, K. Goodman, and L. Xing[a)]
Department of Radiation Oncology, Stanford University School of Medicine, Stanford, California 94305-5847

(Received 9 February 2005; accepted for publication 21 September 2005; published 17 November 2005)

Four-dimensional (4D) CT is useful in many clinical situations, where detailed abdominal and thoracic imaging is needed over the course of the respiratory cycle. However, it usually delivers a larger radiation dose than the standard three-dimensional (3D) CT, since multiple scans at each couch position are required in order to provide the temporal information. Our purpose in this work is to develop a method to perform 4D CT scans at relatively low current, hence reducing the radiation exposure of the patients. To deal with the increased statistical noise caused by the low current, we proposed a novel 4D penalized weighted least square (4D-PWLS) smoothing method, which can incorporate both spatial and phase information. The 4D images at different phases were registered to the same phase via a deformable model, thereby, a regularization term combining temporal and spatial neighbors can be designed for the 4D-PWLS objective function. The proposed method was tested with phantom experiments and a patient study, and superior noise suppression and resolution preservation were observed. A quantitative evaluation of the benefit of the proposed method to 4D radiotherapy and 4D PET/CT imaging are under investigation. © *2005 American Association of Physicists in Medicine*. [DOI: 10.1118/1.2122567]

Key words: 4D CT, low dose, 4D radiotherapy planning, PET/CT, PWLS, deformable registration

I. INTRODUCTION

In radiation therapy, respiratory motion poses significant challenges for tumors in the thorax or abdomen. It can distort the shape of an object, degrade the anatomic position reproducibility during imaging, and necessitate larger margins during radiotherapy planning. It also causes inaccuracy in estimating the tumor volume, thereby preventing an effective dose escalation for the treatment of a target tumor. These issues make it difficult to achieve the desired goals of conformal radiotherapy. Four-dimensional (4D) CT scans, acquired synchronously with a respiratory signal, provide not only the three-dimensional (3D) spatial information, but also temporal changes of the anatomy as a function of the respiratory phase during the imaging, and can therefore be employed in 4D treatment planning to explicitly account for the respiratory motion.[1,2]

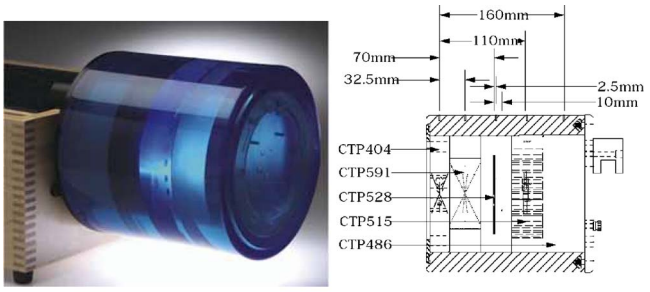
To acquire 4D-CT scans, a position-monitoring system used for respiratory motion tracking is usually interfaced with the CT scanner, so that the CT data are acquired in correlation with real-time positioning. One scanning protocol, which has recently been developed to achieve 4D-CT imaging with the multislice CT scanner,[3,4] uses a "step-and-shoot" technique (cine mode), in which CT projections are acquired over a complete respiratory cycle at each couch position. The period of each CT acquisition segment is time stamped with an "x-ray ON" signal and recorded by the tracking system. The 4D-CT data are subsequently sorted into groups according to their phases of the breathing cycle. Typical parameters for thoracic imaging are 0.45 s cine intervals for a duration slightly longer (1 s) than a full respiratory cycle at each couch position, a 0.5 s gantry rotation, 140 kVp, 175 mA, a 2.5 mm slice thickness, and 10 mm couch increments for a four-row scanner.[5] This data acquisition technique takes, on average, about one minute, depending on the patient breathing period and axial dimension of the scan range.

When a modern multislice CT is used for a regular clinical exam, the dose received by the patient may approach 10 mSv for the head, and 20 mSv for the chest or abdomen. With a 4D acquisition, since a patient is scanned multiple times at each couch position during the imaging, the radiation exposure will be considerably higher than the regular CT scan (up to one order of magnitude higher). The effective dose reduction is thus highly desirable for a clinical application of the cutting-edge 4D-CT scanning technology. In this work we develop a novel method to perform the 4D scan at a lower current and retrospectively generate images comparable to a high-mA 4D scan, hence reducing the patient radiation dose. The central idea is to map different phases into a particular phase through deformable model registration methods, and to improve the image for this phase by statistical estimation from the registered images. In the following, we describe the method and the results of phantom and patient studies in detail.

II. METHODS AND MATERIALS

A. Phantom data acquisition

Two phantom experiments were carried out in this work: one is a commercial calibration phantom CatPhan® 600 (The Phantom Laboratory, Inc., Salem, NY), as shown in Fig. 1, and the other is an anthropomorphic thorax phantom (Fig. 2).



Module	Distance from section 1 center
CTP404, HU calibration, spatial linearity	0mm
CTP591, Bead geometry	32.5mm
CTP528, 21 line pair high resolution	70mm
CTP528. Point source	80mm
CTP515, Subslice and supra-slice low contrast	110mm
CTP486, Solid image uniformity module	150mm

FIG. 1. CatPhan® 600 phantom and its module specification. The module section CTP528 has high resolution multiple line pairs and was used in this work to test the spatial resolution loss during the proposed image processing.

To acquire the 4D CT data, each phantom was placed on the top of a platform capable of sinusoidal motion along the cranial–caudal direction. The amplitude of motion varied discretely from 1 to 6 cm, and the period was continuously adjustable from 0.5 s to 1 min. A Varian Real-time Position Management (RPM) respiratory gating system (Varian Medical Systems, Palo Alto, CA) was used to record the motion by tracking two infrared reflective markers, rigidly mounted on a plastic block on the top of the phantom, by means of an infrared video camera mounted on the PET/CT table (see Fig. 3). The motion amplitude was displayed as a function of time at a rate of 30 Hz on the RPM workstation, and the data were recorded for the entire 4D-CT scan. During the scan, the RPM system also recorded the state of a TTL (Transistor-Transistor Logic) "x-ray ON" signal from the CT scanner, indicating the acquisition time of the CT data, thereby time stamping each CT slice with respiratory motion. Figure 4 shows one example of the motion wave recorded by the

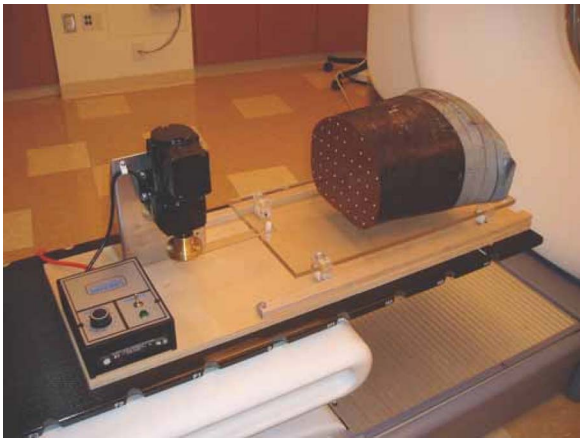


FIG. 2. The experimental design showing the power supply that drives the electric motor, which in turn moves the motion platform and phantom sinusoidally in the cranial–caudal direction as it moves through the CT bore.

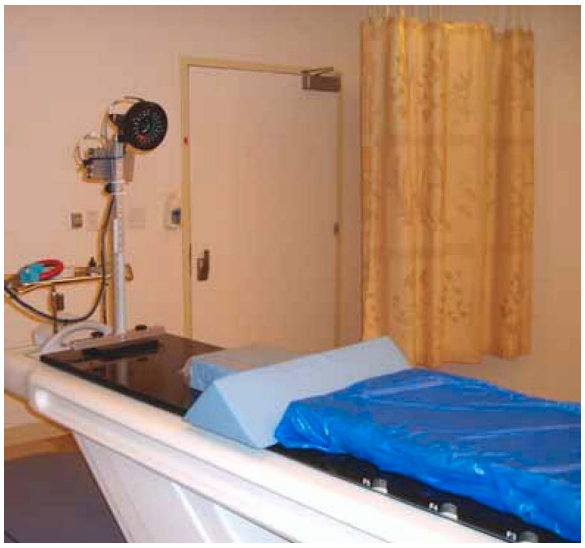


FIG. 3. The RPM infrared camera and illuminator system is mounted at the foot of the couch and used to track the motion of infrared reflectors placed on the top of the thorax phantom or the diaphragm of the patient.

RPM system, in which the diamond markers indicate the motion status when a CT set is generated at each couch position.

A Discovery ST PET/CT scanner (Discovery ST/LightSpeed 8-slice; General Electric Medical Systems) was used in this study. For CT scans, it has a 50 cm transaxial field of view and an available slice thickness of between 0.625 and 10.0 mm. The tube current can be varied between 10 and 440 mA, and the available tube voltage settings are 80, 100, 120, and 140 kVp. The detector coverage in the cranial–caudal direction is 2 cm, and the minimum gantry rotation time is 0.5 s.

In our studies, both phantoms moved with a period of 5.0 s and amplitude of 2 cm. The CT data were acquired in cine mode, at 120 kVp, 10 mA, with cine CT axial field of view of 20 mm (8×2.5 mm slice thickness). The x-ray tube angular velocity was set to 0.5 s/rotation. The "shoot" period at each couch position (cine duration) was set slightly longer than the motion period to 6.1 s, and the cine interval between images was 0.45 s. Each image was reconstructed with 360° of data. The scan covered 10 cm (5 couch positions) with the total acquisition time of about 30 s. For CatPhan® 600, the image can be compared with the specification of the phantom manufacture; to compare the image quality for the thorax phantom, another 4D scan was repeated at 100 mA with all other parameters kept the same.

B. Patient data acquisition

A clinical 4D-CT patient study was performed on the same scanner. The patient was asked to breathe normally during the scan. The plastic block with two infrared reflective markers was taped on the top of the patient's abdomen, placed medially and a few cm inferior to the xiphoid processes. The respiratory signal of the patient from the RPM was recorded and synchronized with CT data acquisition,

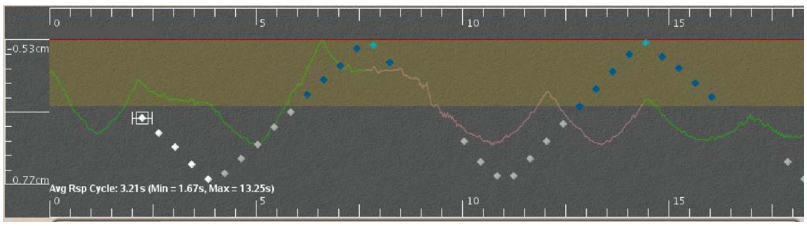


FIG. 4. An illustration of the respiratory motion waveform recorded by the RPM. Each diamond marker reflects the motion status when a CT volumetric image (for one couch position) is generated. The example here shows that 16 CT images were generated in 6.1 s at each couch position.

similar to the phantom scan. The axial coverage of the patient was 25 cm. Other scan parameters were 90 mA, 120 kV, and 2.5 mm slice thickness. The dose to the patient from the 4D-CT was 80.2 mGy. The cine interval between images was at 0.45 s. Each image reconstruction used 360° of data corresponding to 0.8 s duration.

C. Deformable registration for phase images

After the scan data were prospectively reconstructed at the PET/CT scanner, both the CT images and the corresponding motion data recorded by the RPM system were transferred to a GE Advantage Workstation (GE Medical Systems, Waukesha, WI). The "Advantage 4D" software on the workstation simultaneously displays the CT images and the motion data, and sorts the cine images into a set of respiratory phase images. For both the phantom and the patient scans in this study, a total of 10 phases were created, with phase intervals of 10% of the respiratory cycle.

The phase images obtained from low-mA scans are quite noisy (see examples in Figs. 6 and 8, later) and usually cannot be used directly for 4D radiation treatment planning. Before applying our 4D image enhancement technique to these noisy phase images, a registration step was performed in order for the temporal information to be appropriately incorporated in the next step. To account for the complex organ motion, we have previously investigated a few deformable models,[6,7] and adopted the free-form spline model (BSpline) in this work.[8] Its simplicity and yet accuracy make it a preferred tool for clinical applications.[9–11] In this model, a lattice of user-defined nodes is overlaid on the image. Each node contains a deformation vector, whose components are to be determined by the optimization procedure. The deformation at any point of the image is calculated by spline interpolation of closest nodes values. Unlike other spline models, the BSplines are locally controlled, such that the displacement of an interpolation point is influenced only by the closest grid points and changing a lattice node only affects the transformation regionally, making it efficient in describing local deformations. Suitable node deformations are solved using the gradient-based algorithm L-BFGS[7,12] due to its superior performance in large-scale optimization problems. The optimizer iteratively varies deformation values to minimize the metric, a mathematical measure of similarity between images. The normalized cross correlation (NCC) metric (described in the next section) was used here since the registration was applied on images acquired under identical settings.[13]

To improve the 4D images, for example, at phase i, all other phases were registered to phase i and resulted in transformations representing a temporal sequence of 3D deformation fields, or in other words, a 4D model of organ motion. Each phase throughout the respiratory cycle can be subsequently warped with the deformation field to map the points in each phase to the corresponding points in the reference image, i.e., the phase i image.

D. 4D-CT image enhancement

The goal of image enhancement is essentially to obtain a more accurate estimation of the "true" intensity (CT number) of each voxel in the image being processed. In our case, the information available here included the image at phase i, and nine other images that were previously registered to phase i. Ideally, if the nine phases are perfectly registered to phase i, then each of them can be considered a repeated scan for phase i. Therefore, it is natural to average them with phase i to get the improved image. In reality, however, there are always errors in registration, and not all voxels will be able to match exactly. Simple averaging will thus lead to blurring in the image. In the following, we describe a statistical method to incorporate the available information into a *penalized weighted least square* (PWLS) objective function to achieve the optimal estimation of the true image at phase i. The method extends the conventional PWLS method[14,15] into four dimensions to include the temporal information, and from here on we refer to it as the 4D-PWLS smoothing method.

Let $\mathbf{X}$ denote the "4D deformed volumetric image" $\mathbf{X} = (x_1, x_2, \ldots, x_N, x_{N+1}, x_{N+2}, \ldots, x_{2N}, \ldots, x_{(M-1)N+1}, x_{(M-1)N+2}, \ldots, x_{MN})'$, where M is the total number of phases after deformable registrations, N is the total number of voxels in each phase image, and the prime ' ' denotes the transpose operation. Let $\boldsymbol{\mu}$ be the unknown true 4D image $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_N, \mu_{N+1}, \mu_{N+2}, \ldots, \mu_{2N}, \ldots, \mu_{(M-1)N+1}, \mu_{(M-1)N+2}, \ldots, \mu_{MN})'$. The 4D-PWLS solution is to find $\hat{\boldsymbol{\mu}}$ that minimizes the following objective function:

$$\Phi(\boldsymbol{\mu}) = \frac{1}{2}(\mathbf{X} - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{X} - \boldsymbol{\mu}) + \alpha \mathbf{H}(\boldsymbol{\mu})$$

$$\hat{\boldsymbol{\mu}} = \arg\min \Phi(\boldsymbol{\mu}), \tag{1}$$

where the first term is the weighted-least-square part, in which the variance–covariance matrix Σ is diagonal with the jth entry σ_j^2, $j = 1, 2, \ldots, N \times M$, if the intensity measurement at each voxel of the 4D image is statistically independent. To

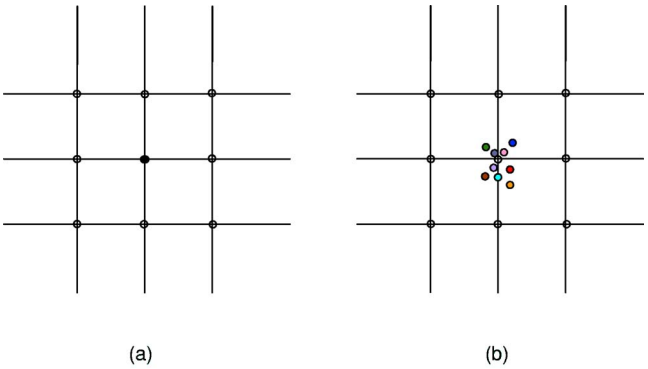


FIG. 5. An illustration of the registration error. (a) A perfect match of the corresponding voxels; (b) when errors exist, the voxels transformed by registration may not match exactly their targeted locations.

provide a relatively accurate estimation of the variance σ_j^2 at each voxel j, a practical way is to select a "window" for each voxel and calculate the sample variance in the window as the estimation. For 4D CT, the window can be selected as a set that includes neighbors in both the spatial and temporal domains, for example, the four nearest spatial neighbors within frame i, and $(M-1)$ neighbors at the corresponding locations in the registered frames.

The second term in (1) is the regularizing penalty term that encourages the neighborhood smoothness, and the coefficient $\alpha>0$ is a constant to weight the penalty.[16,17] Because both spatial and temporal neighbors are involved, in this work, a simple quadratic form was selected to penalize the two kinds of neighbors:

$$\mathbf{H}(\boldsymbol{\mu}) = \frac{1}{2}\sum_j \sum_{k\in N_j} w_{jk}\frac{1}{2}(\mu_j - \mu_k)^2, \tag{2}$$

where N_j is the set of "4D neighbors" of the jth pixel, which include eight in-plane spatial neighbors and nine temporal (phase) neighbors in our studies. As illustrated in Fig. 5, if each phase is perfectly registered to a reference image (for example, phase i), then the corresponding voxels overlap with each other (Fig. 5(a)); when errors are present in registration (due to noise, deformable model, interpolation, etc.), the points in the registered image may not be transformed to the corresponding point in the reference image, but some place nearby, as shown in Fig. 5(b). Therefore, these corresponding voxels in the registered phases can be considered as the new neighbors of the reference image of phase i, in addition to its own 3D spatial neighbors. The weights w_{jk} equal 1 for horizontal and vertical spatial neighbors, and $1/\sqrt{2}$ for diagonal neighbors. For the temporal neighbors, the weights were proportional to the NCC between current frame m and frame i, which is defined as

$$f_{im} = \frac{\sum_{j=1}^{N} x_{j+(m-1)N}x_{j+(i-1)N}}{\sqrt{\sum_{j=1}^{N}(x_{j+(m-1)N})^2\sum_{j=1}^{N}(x_{j+(i-1)N})^2}}. \tag{3}$$

The proportion is determined by the ratio of the NCCs in temporal and spatial directions. Specifically, the NCCs of

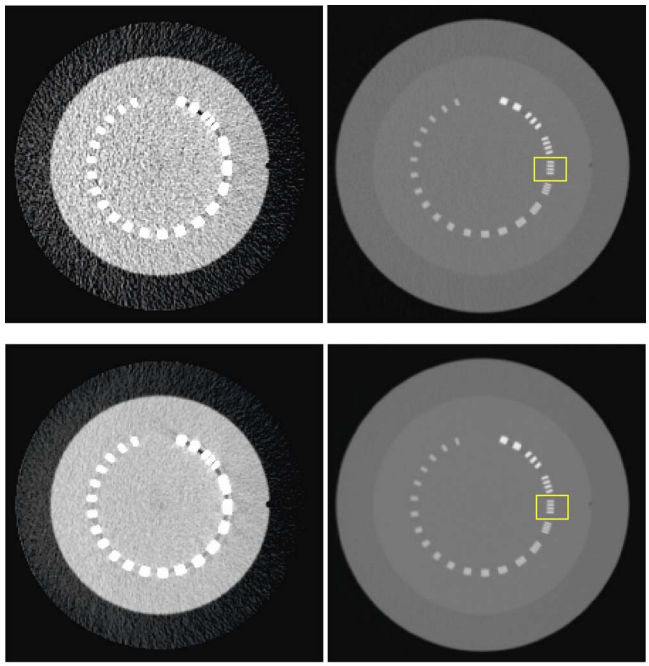


FIG. 6. Phantom study for the 4D-PWLS method with the CatPhan® 600 phantom. The CTP528 section (high-resolution line pairs) is compared in two window width/level settings for 10 mA CT image before (top row) and after (bottom row) the proposed 4D-PWLS smoothing. The left column is displayed with window width 200, level 80; the right column is displayed with window width 1500, level 150.

adjacent rows in frame i is calculated similarly as (3), and an averaged similarity metric $\bar{f}_s$ is obtained for frame i for the spatial direction. Naively, larger the similarity measure is (maximum 1), higher the weight should be. Therefore, we set the weights of temporal neighbors to be $f_{im}/\bar{f}_s$. Note that for the objective function $\Phi(\boldsymbol{\mu})$ to be convex, the temporal weights must also be positive. A unique simple closed-form solution to (1) exists:

$$\hat{\boldsymbol{\mu}} = (\boldsymbol{\Sigma}^{-1} + \alpha\mathbf{H})^{-1}\boldsymbol{\Sigma}^{-1}\mathbf{X}. \tag{4}$$

Since the matrix $\mathbf{H}$ is multidiagonal, the inversion involved in (4) is not straightforward. In this work we adopt a simple iterative algorithm, known as the *iterated conditional mode* (ICM),[18]

$$\mu_j^{(n+1)} = \left(x_j + \alpha\sigma_j^2\sum_{k\in N_j} w_{jk}\mu_k^{(n)}\right) \Big/ (1+\alpha\sigma_j^2 w_j), \tag{5}$$

$$w_j = \sum_{k\in N_j} w_{jk}, \tag{6}$$

which converges to the unique solution.

III. RESULTS

A. Phantom study

The CatPhan® 600 phantom is capable of measuring the spatial resolution. In Fig. 6 we compared the images for one slice of the CatPhan® 600 that contains multiple strips on a circle, of which the resolution sections are ranging from 1 to

TABLE I. Mean and standard deviation (SD) of the CT number in the selected ROIs before and after the proposed 4D smoothing for the 10 mA scan of the CatPhan 600 phantom. ROI-1 is the 10×10 pixel square region in the center, ROI-2 is the 10×10 pixel square region 5.0 cm above the center, ROI-3 is the 10×10 pixel square region 8.5 cm above the center.

	ROI-1	ROI-2	ROI-3
Mean CT number (HU) in original image	94.75	100.19	12.5
Mean CT number after 4D smoothing (HU)	95.94	100.12	11.43
SD CT number (HU) in original image	17.39	15.76	13.12
SD CT number after 4D smoothing (HU)	7.46	6.24	5.31

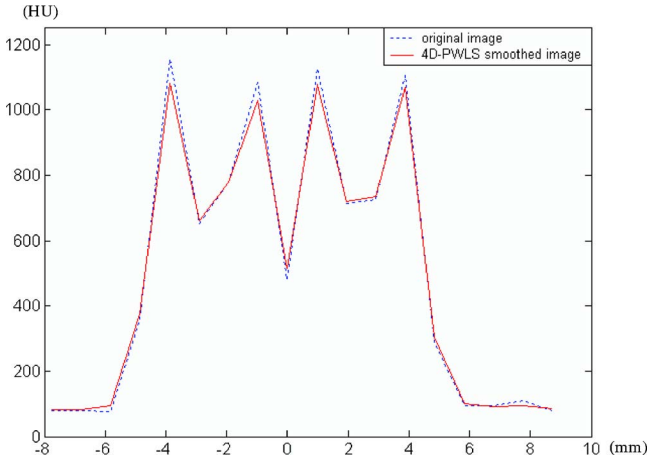


FIG. 7. Profiles across the strips indicated by the squares in the images of Fig. 6. The dotted line shows the profile of the original 10 mA data, and solid line is the profile for the smoothed data.

21 lines pairs per cm. The original and smoothed CT images are compared in two different window width/level settings, in which the left column is displayed at window width 200, level 80, and the right column is displayed at window width 1500, level 150. The top row is the original 4D CT images acquired at 10 mA, and the bottom is the smoothed images with $\alpha=1.0\times10^{-6}$. Since the proposed weighted least square algorithm is adaptive to the local variance at each pixel of the image, the resolution and noise in the smoothed image could be variable at different locations. We examined three regions of interest (ROIs) for measurements of the average CT number and standard deviation, each containing 10×10 pixels in a uniform rectangular region. The first ROI was located at the center of the image, ROI-2 was 5.0 cm above the center, and ROI-3 was 8.5 cm above the center. The results are listed in Table I. It is found that the smoothed image had the similar average CT number as the original image for the measured ROIs, while the standard deviation of the CT numbers were reduced from 17.39, 15.76, 13.12 to 7.46, 6.24, and 5.31, respectively. Thus, about a $2.5\times$ reduction of the variation of CT number was achieved. Obviously, the reduction of noise itself is not enough to assess the overall image quality. We therefore further compared the two sets of images for the spatial resolution. The vertical profiles across the strips in the middle right (see Fig. 6) were plotted in Fig. 7, where only a small resolution loss was observed. The four peak values (or the "signals") from left to right dropped less than 6% from 1155, 1084, 1129, 1105 to 1086.9, 1029.2, 1079.5, 1069.8, respectively, after processing.

In the human thorax phantom study, we compared the smoothed CT images with the original low-mA (10 mA) CT images, as well as the original high-mA (100 mA) CT images. Figure 8 shows their coronal images at five different phases of 0%, 20%, 40%, 60%, and 80%. The left column contains the 4D CT images acquired at 100 mA after sorting by the GE 4D Advantage software. The middle column shows the images at the same phases acquired at 10 mA, and the 4D-PWLS smoothed images with $\alpha=2.0\times10^{-6}$ based on the low-current CT data are displayed in the right column. The effective reduction of the image noise can be observed at a certain cost of the spatial resolution. To quantify this, we

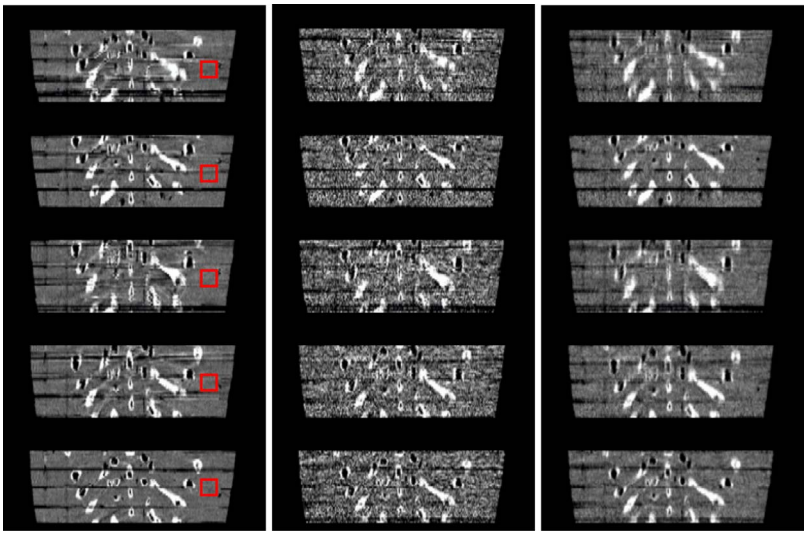


FIG. 8. Phantom study for the 4D-PWLS method with the thorax phantom. The left and middle columns are the original phases obtained from the GE Advantage Workstation, for 100 and 10 mA, respectively; the right column shows the 10 mA phases after 4D-PWLS processing. From top to bottom are phase 0%, 20%, 40%, 60%, 80%, respectively. The red rectangles represent the selected ROI for calculation of SNRs, each of which contains $5\times5\times5$ voxels.

TABLE II. A comparison of SNRs in the thorax phantom study for different phases. The penalty α was 2.0×10^{-6} in the 4D-PWLS smoothing.

	Phase 0%	Phase 20%	Phase 40%	Phase 60%	Phase 80%
100 mA scan	0.479	0.215	0.122	0.226	0.349
10 mA scan	0.078	0.057	0.015	0.045	0.059
PWLS Smoothed 10 mA scan	0.189	0.170	0.086	0.167	0.212

calculated the signal-to-noise ratios (SNRs), defined as MEAN/SD (mean divided by standard deviation) within a selected region of interest shown in Fig. 8, and listed them in Table II. From the table, it is seen that the average SNR of 10 mA images increased by more than three-fold from 0.051 to 0.165 after the proposed post-processing. The average SNR of 100 mA phase images is 0.278.

While useful, the above measure of SNR is not sufficient to characterize the goodness of a smoothing method. To assess the proposed algorithm accurately, we calculated the relative contrast (RC) in addition to the background noise, similar to Ref. 19. In our study, the RC is defined as the signal difference between object and background. Without loss of generality, one slice location of the three datasets at phase 0% was used to measure the RCs and background noises. The slices are shown in Fig. 9 for the three datasets including the high-mA, low-mA, and the smoothed images, from top to bottom, respectively. The object signal was measured as the mean CT number of nine selected holes indicated by the yellow rectangles in Fig. 9, each containing 4×4 pixels; the background signal was the mean CT number of a selected uniform region close to the holes containing 20×20 pixels (the red rectangle in Fig. 9). The RCs and background noise were measured for the proposed algorithm for different settings of the smoothing parameter α, and their relations are plotted in Fig. 10, where the red lines represent values for the 100 mA data, and the diamond marks are for the smoothed data at a different smoothing level. When $\alpha = 0$, it represents the original low-mA (10 mA) data. It is found that the background noise decreases rapidly with α at the beginning and the decreasing rate gets slower when α increases further, while the contrast is reduced almost linearly with α. According to the plots shown in Fig. 10, a good choice for the parameter α in this case seems to be between 1×10^{-6} and 1×10^{-5} because of a large percentage of reduction of the noise and relatively small loss of the contrast. In Table III, we list the measured RC and noise for the proposed 4D PWLS method with an "optimal" α empirically chosen as 2×10^{-6}. Compared with the original low-mA data, the background noise was reduced approximately 2.7 times, with only 1.3% loss of the contrast.

B. Patient study

The patient study for the end-inspiration phase (phase 0%) and end-expiration phase (phase 50%) are shown in Figs. 11(a) and 12(a), respectively. The left column in each figure is the original phase image obtained from GE Advantage Workstation, and the right column is the image after processing with the proposed 4D-PWLS enhancement. From top to bottom are the axial, coronal, and sagittal displays, respectively. Successful noise suppression is observed for both phases in these pictures. With $\alpha = 1.0 \times 10^{-5}$ the SNRs increased from 2.204 to 4.558 for the end-expiration phase, and from 1.741 to 3.862 for the end-inspiration phase in the selected ROIs (red rectangles in the figures, each consisting of 125 voxels). In Figs. 11(b) and 12(b), the horizontal profiles across the center of the transaxial slices are compared for the original and processed images. It is found that the 4D-PWLS processing preserves the CT values and edge information very well while reducing the noise. This patient study was performed using a 4D protocol using x-ray tube current of 90 mA. For a thorax scan, it is possible that the current can be reduced. However, it is important to emphasize that the proposed technique represents an independent way to reduce the patient radiation dose.

In order to demonstrate the role of the temporal information in suppressing noise, we compared the same algorithm with and without the aid of the registered phases. By assigning the weights of the temporal neighbors to zero, a conventional 3D-PWLS algorithm can be obtained, which smoothes an image based only on the 3D spatial correlation. Results of the end-expiration phase using the 3D PWLS method with same α are shown in the top row of Fig. 13. The SNRs for the selected ROI increase from 2.204 to 2.783, compared to 4.558 of the 4D-PWLS method. Furthermore, for a qualitative examination of the spatial resolution, we compared 3D- and 4D-PWLS results by subtracting from each of them the original noisy image. The resulting difference images are shown in the middle and bottom rows in Fig. 13, where better edge preservation is observed for the 4D-PWLS method as well.

IV. DISCUSSION AND CONCLUSION

We have proposed a 4D-PWLS smoothing method to reduce the noise in 4D-CT images. The technique allows us to obtain high quality 4D-CT images based on the data acquired with a low x-ray tube current at individual phases, resulting in a significant reduction in the patient radiation dose. Through deformable registration, the method incorporates the information of different phases into one objective function and finds the optimal estimation of the true image in terms of the least square metric based on the first- and second-order statistics of the data. In this method, the data in the temporal domain are incorporated into the penalty term

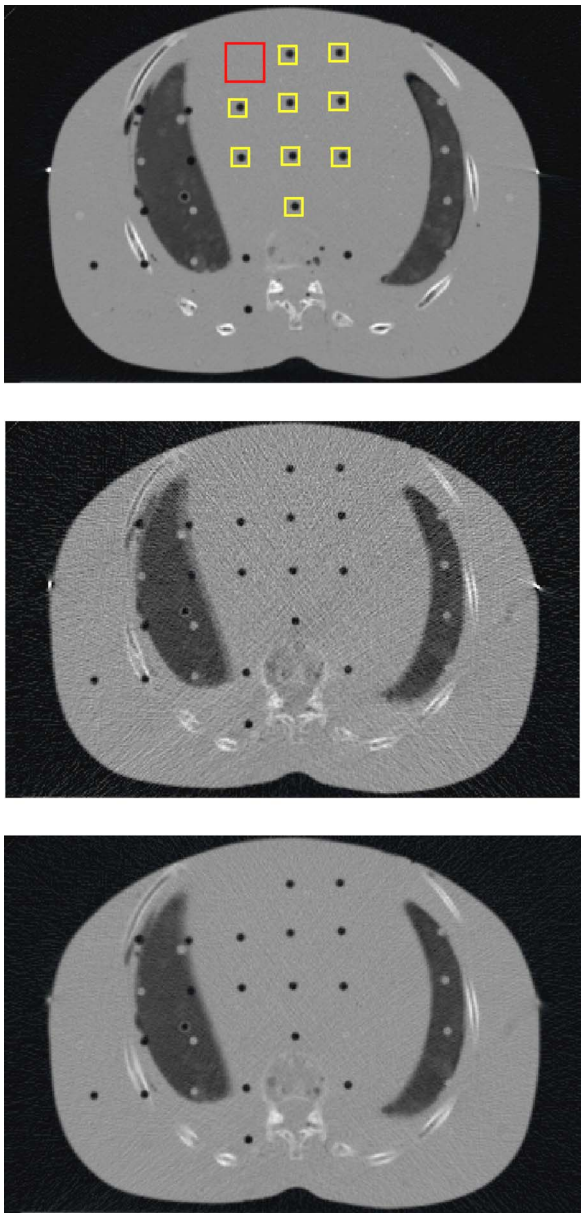


FIG. 9. Axial CT images of the thorax phantom. From top to bottom are the 100 mA, 10 mA, and 4D-PWLS smoothed 10 mA images. The yellow squares show the air holes for measuring the "signal," the red square shows the region for measuring the "background." The relative contrast is defined as the CT number difference between the signal and the background.

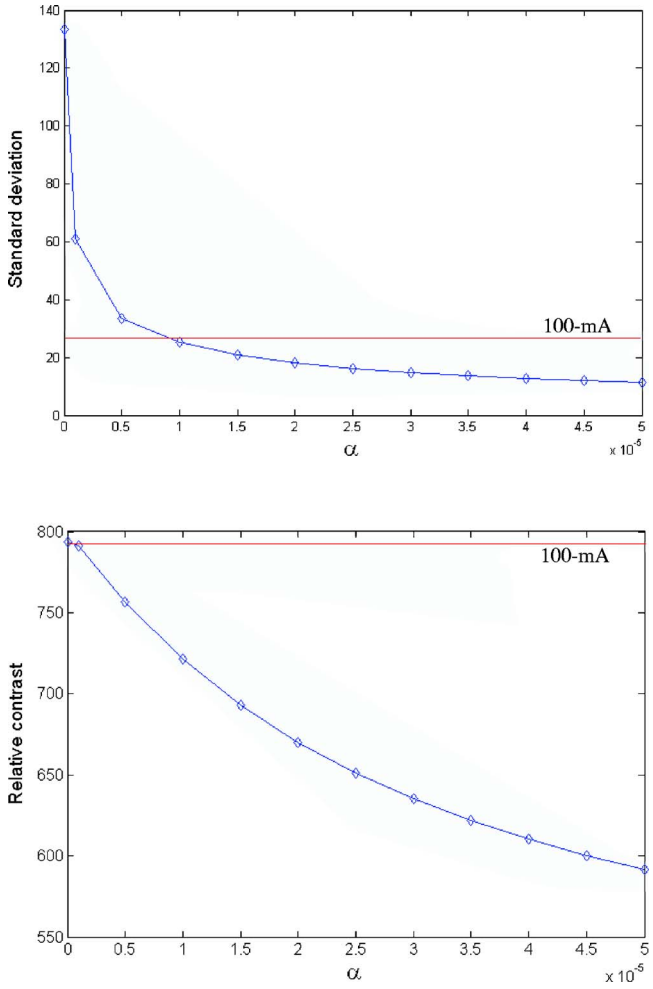


FIG. 10. The relations of noise and relative contrast with the penalty weight α. The noise is characterized by the standard deviation of the CT numbers in the uniform region shown in Fig. 9. The red lines in the figure show the noise and relative contrast of the 100 mA image. The blue line and diamond marks are for the smoothed images with different α. When $\alpha=0$ it represents the 10 mA original data.

as additional neighbors, and their "distances" are determined by the maximized NCC in the registration step to accommodate possible local inaccuracies. One concern of this method is that if the tumor contour is compromised after the deformation because of the existence of small uncertainty in deformable registration. The experience from another group[20] and our group[21] have indicated that generally an accuracy of less than 2–3 mm is achievable, and the statistical averaging strategy as described above should further reduce the uncertainty. While the method seems to be very robust, better registration will definitely improve the performance of the proposed technique.

The proposed 4D-PWLS technique has been validated by phantom experiments and applied to patient data as well, where it is observed that the method effectively smoothes out the noise and leads to clear improvements of the image quality. Basically, the method integrates ten low-dose phases into one phase, hence will result in an image comparable to a ten-times higher dose scan in an ideal situation (perfect registration), which means a maximum of ten-fold dose could be saved. Note that the 4D PWLS is essentially a low-pass

TABLE III. A comparison of relative contrast and noise in the thorax phantom study.

	100 mA scan	10 mA scan	Smoothed 10 mA scan
Relative contrast (HU)	793.60	793.72	783.95
Background noise (HU)	26.98	133.48	48.86
Contrast to noise ratio	29.4	5.9	16.0

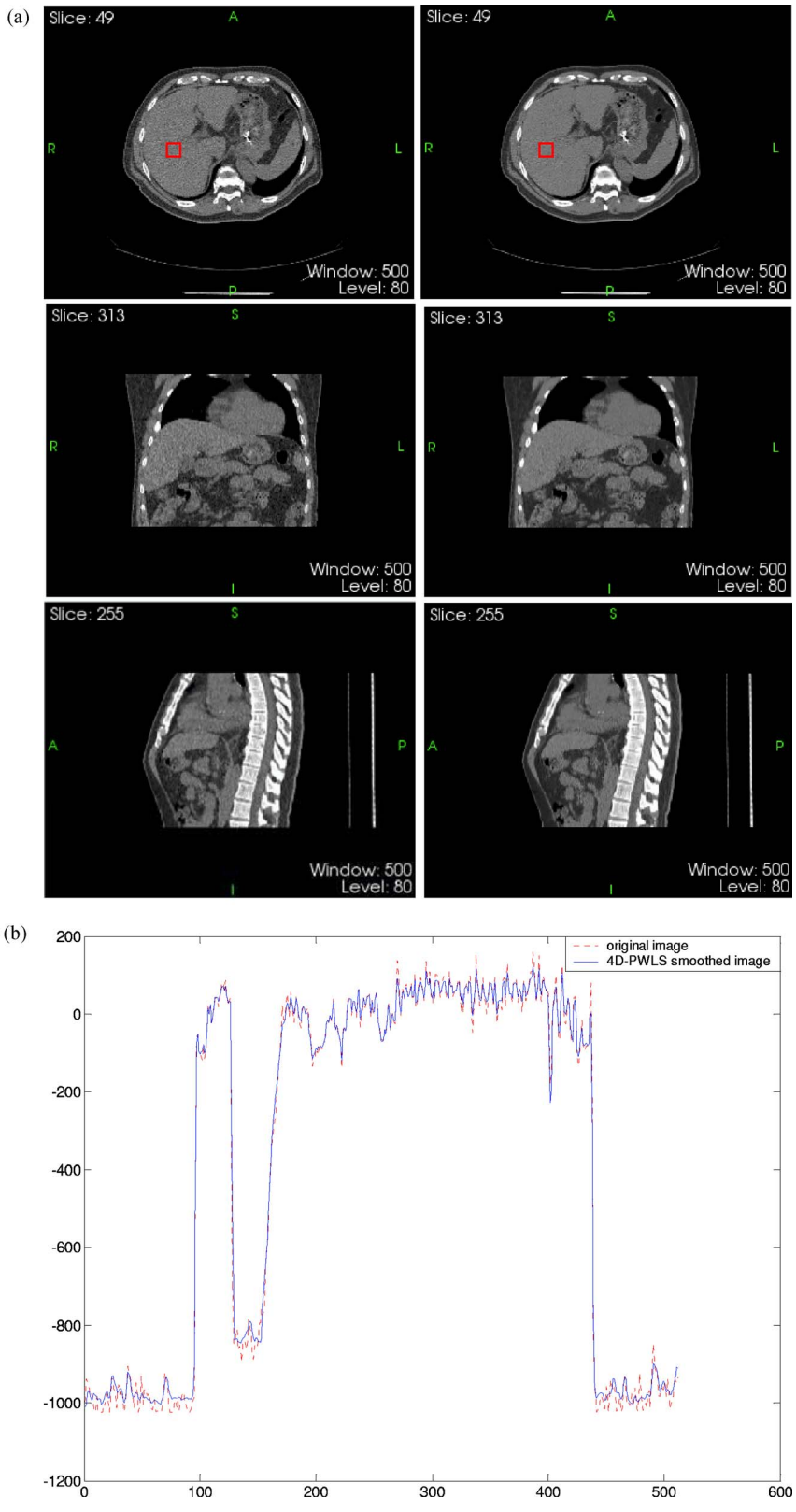


FIG. 11. (a) Patient study for the 4D-PWLS method at the end-inspiration phase. The left column contains the original images acquired from the GE Advantage Workstation, and the right column shows the image after 4D-PWLS processing. The red rectangles represent the selected ROI for the calculation of SNRs, each of which contain $5\times5\times5$ voxels. (b). A comparison of the horizontal profiles across the center of the transaxial images in (a).

filtering method, which inevitably reduces the image spatial/temporal resolution while smoothing the noise. The tradeoff between the noise and resolution can be controlled by selecting an appropriate value for the penalty weight α, so that a reasonably enhanced image can be achieved, as demonstrated in this work. However, the way to determine the optimal value for the parameter α is yet empirical. The overall evaluation of the smoothing method eventually depends on the performance tests of tumor detectability. For example, to quantify the effectiveness of the method in terms of the clini-

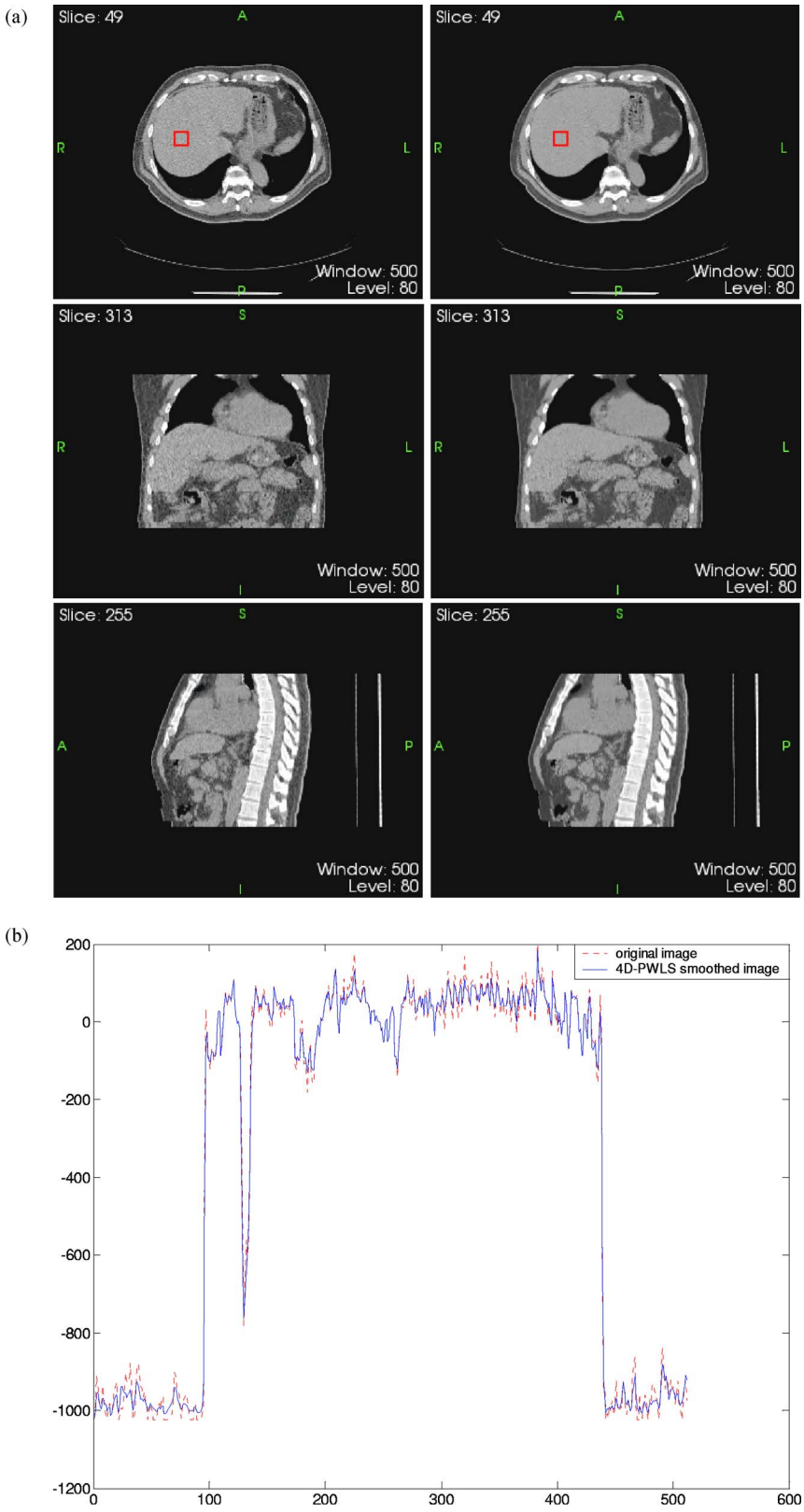


FIG. 12. (a) Patient study for the 4D-PWLS method at the end-expiration phase. The left column contains the original images acquired from the GE Advantage Workstation, and the right column shows the image after 4D-PWLS processing. The red rectangles represent the selected ROI for calculation of SNRs, each of which contain $5\times5\times5$ voxels. (b). A comparison of the horizontal profiles across the center of the transaxial images in (a).

cal goal in practice, *receiver operating characteristic* (ROC) studies should be carried out with large amount of data,[22] which is beyond the scope of the current work. Our main purpose in this paper is to present a statistic framework to incorporate temporal information into image restoration. The ICM algorithm in our studies usually converged to a satisfied solution within ten iterations. However, because of a quadratic prior being used, other algorithms such as a conjugate gradient may be developed for even better convergence. Furthermore, it may be of research interest to develop a penalty

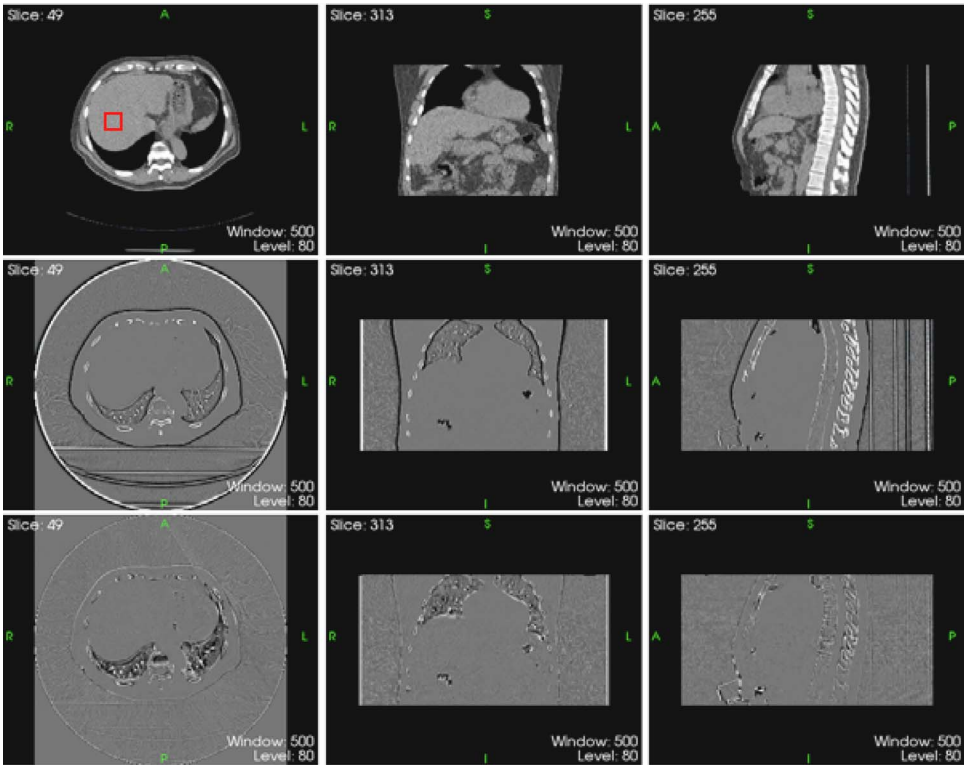


FIG. 13. A comparison of PWLS with and without registered phase information. The top row is the result of 3D PWLS smoothing without using temporal information for the end-inspiration phase patient data. The middle row shows the difference images between the 3D-PWLS smoothing results (top row in this figure) and original images [the left column in Fig. 11(a)]. The bottom row contains the difference images between 4D-PWLS smoothing results [the right column in Fig. 11(a)] and original images. The fewer edges observed in the difference image indicate an improved preservation of the spatial resolution in the 4D method. The red rectangles represent the selected ROI for calculation of SNRs, each of which contain $5\times5\times5$ voxels.

term other than the quadratic form for better preservation of the spatial/temporal resolution for 4D CT.

In our earlier study,[15] it was found that further improvements to image quality may be possible if the noise reduction is performed in sinogram space (before reconstruction) rather than in image space (after reconstruction). In future research, we will study the possibility of applying deformable registration and 4D-PWLS to sinogram space. A quantitative evaluation of the benefits of the proposed method to applications in PET/CT imaging[23–25] and 4D treatment planning for radiotherapy[26,27] are also under investigation.

ACKNOWLEDGMENTS

This research is supported in part by the Department of Defense (DAMD17-03-1-0657) and the National Cancer Institute (5 R01 CA98523-01).

[a]Author to whom correspondence should be addressed. Department of Radiation Oncology, Stanford University School of Medicine, Clinical Cancer Center, 875 Blake Wilbur Drive, Rm CC-G204, Stanford, California 94305-5847. Telephone: (650) 498-7896; Fax: (650) 498-4015; Electronic mail: lei@reyes.stanford.edu

[1]S. S. Vedam, P. J. Keall, V. R. Kini, H. Mosafavi, H. P. Shukla, and R. Mohan, "Acquiring a four-dimensional computed tomography dataset using an external respiratory signal," Phys. Med. Biol. **48**, 45–62 (2003).

[2]P. J. Keall, G. Starkschall, H. Shukla, K. M. Forster, V. Ortiz, C. W. Stevens, S. S. Vedam, R. George, T. Guerrero, and R. Mohan, "Acquiring 4D thoracic CT scans using a multislice helical method," Phys. Med. Biol. **49**, 2053–2067 (2004).

[3]D. A. Low, M. Nystrom, E. Kalinin *et al.*, "A method for the reconstruction of four-dimensional synchronized CT scans acquired during free breathing," Med. Phys. **30**, 1254–1263 (2003).

[4]T. Pan, T.-Y. Lee, E. Rietzel, and G. T. Y. Chen, "4D-CT imaging of a volume influenced by respiratory motion on multi-slice CT," Med. Phys. **31**, 333–340 (2004).

[5]Y. E. Erdi, S. A. Nehmeh, T. Pan, S. Kohlmyer, O. D. Squire, H. Schoder, H. W. Yeung, J. L. Humm, K. E. Rosenzweig, and S. M. Larson, "The CT motion quantitation of lung lesions and its impact on PET measured SUV's,"J. Nucl. Med. **45**, 1287–1292 (2004).

[6]J. Lian, L. Xing, S. Hunjan, C. Dumoulin, J. Levin, A. Lo, R. Watkins, K. Rohling, R. Giaquinto, D. Kim, D. Spielman, and B. Daniel, "Mapping of the prostate in endorectal coil-based MRI/MRSI and CT: A deformable registration and validation study," Med. Phys. **31**, 3087–3094 (2004).

[7]E. Schreibmann and L. Xing, "Narrow band deformable registration of prostate MRI/MRSI and CT studies," Int. J. Radiat. Oncol., Biol., Phys., **62**, 595–605 (2005).

[8]D. Mattes, D. R. Haynor, H. Vesselle, T. K. Lewellen, and W. Eubank, "PET-CT image registration in the chest using free-form deformations," IEEE Trans. Med. Imaging **22**, 120–128 (2003).

[9]T. Rohlfing, C. R. Maurer, Jr., D. A. Bluemke, and M. A. Jacobs, "Volume-preserving nonrigid registration of MR breast images using free-form deformation with an incompressibility constraint," IEEE Trans. Med. Imaging **22**, 730–741 (2003).

[10]D. Rueckert, L. I. Sonoda, C. Hayes, D. L. Hill, M. O. Leach, and D. J. Hawkes, "Nonrigid registration using free-form deformations: Application to breast MR images," IEEE Trans. Med. Imaging **18**, 712–721 (1999).

[11]D. Loeckx, F. Maes, D. Vandermeulen, and P. Suetens, "Non-rigid image registration using a statistical spline deformation model," Info. Process. Med. Imaging **18**, 463–474 (2003).

[12]D. C. Liu and J. Nocedal, "On the limited memory BFGS method for large scale optimization," Math. Program. **45**, 503–528 (1989).

[13]B. F. Hutton and M. Braun, "Software for image registration: Algorithm, accuracy, efficacy," Semin. Nucl. Med. **33**, 180–192 (2003).

[14]J. A. Fessler, "Penalized weighted least-squares image reconstruction for positron emission tomography," IEEE Trans. Med. Imaging **13**, 290–300 (1994).

[15]T. Li, X. Li, J. Wang, J. Wen, H. Lu, J. Hsieh, and Z. Liang, "Nonlinear sinogram smoothing for low-dose x-ray CT," IEEE Trans. Nucl. Sci. **51**, 2505—2513 (2004).

[16]S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," IEEE Trans. Pattern Anal. Mach. Intell. **6**, 721–741 (1984).

[17]X. Li, L. Li, H. Lu, and Z. Liang, "Partial volume segmentation of brain magnetic resonance images based on maximum *a posteriori* probability,"

Med. Phys. **32**, 2337–2345 (2005).
[18]J. Besag, "On the statistical analysis of dirty pictures," J. R. Stat. Soc. Ser. B. Methodol. **48**, 259–302 (1986).
[19]P. K. Saha and J. K. Udupa, "Scale-based diffusive image filtering preserving boundary sharpness and fine structures," IEEE Trans. Med. Imaging **20**, 1140–1155 (2001).
[20]M. M. Coselmon, J. M. Balter, D. L. McShan, and M. L. Kessler, "Mutual information based CT registration of the lung at exhale and inhale breathing states using thin-plate splines," Med. Phys. **31**, 2942–2948 (2004).
[21]E. Schreibmann, G. T. Y. Chen, and L. Xing, "Image interpolation in 4D CT using a BSpline deformable registration model," Int. J. Radiat. Oncol., Biol., Phys..
[22]T. Li, J. Wen, G. Han, H. Lu, and Z. Liang, "Evaluation of an efficient compensation method for quantitative fan-beam brain SPECT reconstruction," IEEE Trans. Med. Imaging **24**, 170–179 (2005).
[23]S. A. Nehmeh, Y. E. Erdi, T. Pan, A. Pevsner, K. E. Rosenzweig, E. Yorke, G. S. Mageras, H. Schoder, Phil Vernon, O. Squire, H. Mostafavi, S. M. Larson, and J. L. Humm, "Four-dimensional (4D) PET/CT imaging of the thorax," Med. Phys. **31**, 3179–3186 (2004).
[24]S. A. Nehmeh, Y. E. Erdi, K. E. Rosenzweig, H. Schoder, S. M. Larson, O. D. Squire, and J. L. Humm, "Reduction of respiratory motion artifacts in PET imaging of lung cancer by respiratory correlated dynamic PET: Methodology and comparison with respiratory gated PET," J. Nucl. Med. **44**, 1644–1648 (2003).
[25]S. A. Nehmeh, Y. E. Erdi, T. Pan, E. Yorke, G. S. Mageras, K. E. Rosenzweig, H. Schoder, H. Mostafavi, O. D. Squire, A. Pevsner, S. M. Larson, and J. L. Humm, "Quantitation of respiratory motion during 4D-PET/CT acquisition," Med. Phys. **31**, 1333–1338 (2004).
[26]P. J. Keall, S. Vedam, V. Kini, S. Joshi, G. Tracton, and R. Mohan, "4D IMRT: Planning methodology," Int. J. Radiat. Oncol., Biol., Phys. **54**, S318–S319 (2002).
[27]H. Shirato, S. Shimizu, K. Kitamura *et al.*, "Four-dimensional treatment planning and fluoroscopic real-time tumor tracking radiotherapy for moving tumor," Int. J. Radiat. Oncol., Biol., Phys. **48**, 435–442 (2000).

INSTITUTE OF PHYSICS PUBLISHING PHYSICS IN MEDICINE AND BIOLOGY

Phys. Med. Biol. **51** (2006) 253–267 doi:10.1088/0031-9155/51/2/005

Motion correction for improved target localization with on-board cone-beam computed tomography

T Li, E Schreibmann, Y Yang and L Xing

Department of Radiation Oncology, Stanford University School of Medicine, Stanford, CA 94305-5847, USA

E-mail: lei@reyes.stanford.edu

Received 5 July 2005, in final form 10 October 2005
Published 21 December 2005
Online at stacks.iop.org/PMB/51/253

Abstract

On-board imager (OBI) based cone-beam computed tomography (CBCT) has become available in radiotherapy clinics to accurately identify the target in the treatment position. However, due to the relatively slow gantry rotation (typically about 60 s for a full 360° scan) in acquiring the CBCT projection data, the patient's respiratory motion causes serious problems such as blurring, doubling, streaking and distortion in the reconstructed images, which heavily degrade the image quality and the target localization. In this work, we present a motion compensation method for slow-rotating CBCT scans by incorporating into image reconstruction a *patient-specific* motion model, which is derived from previously obtained four-dimensional (4D) treatment planning CT images of the same patient via deformable registration. The registration of the 4D CT phases results in transformations representing a temporal sequence of three-dimensional (3D) deformation fields, or in other words, a 4D model of organ motion. The algorithm was developed heuristically in two-dimensional (2D) parallel-beam geometry and extended to 3D cone-beam geometry. By simulations with digital phantoms capable of translational motion and other complex motion, we demonstrated that the algorithm can reduce the motion artefacts locally, and restore the tumour size and shape, which may thereby improve the accuracy of target localization and patient positioning when CBCT is used as the treatment guidance.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

A new technology of cone-beam computed tomography (CBCT) has recently been integrated *on-board* with the linear accelerator (linac) in radiotherapy clinics. Superior to the common approaches based on two orthogonal images provided by the mega-voltage electronic

portal imaging device (EPID), CBCT can provide high-resolution three-dimensional (3D) information of the patient anatomy in the treatment position, and thus has great potential for improved target localization and irradiation dose verification in radiotherapy (Jaffray *et al* 1999, Moseley *et al* 2004, Sidhu *et al* 2003, van Herk *et al* 2004), and can also be utilized in synchronized respiratory gating radiotherapy (Jin and Yin 2005). However, when the on-board CBCT is used in imaging the thorax or abdomen of a patient, respiration induced artefacts such as blurring, doubling, streaking and distortion are observed, which heavily degrade the image quality, and affect the target localization ability, as well as the accuracy of dose verification (Sonke *et al* 2005). These artefacts are much more severe than those found in conventional CT examinations. In conventional CT, each rotation of the scan can be completed within 1 s, during this period the organ/tumour motion is relatively small. Furthermore, patient body-restraints and breath-hold techniques can be used to minimize the motion if necessary. In contrast, in a CBCT scan, the gantry rotation speed is much slower, typically 40 s to 1 min for a full 360° scan in acquiring the projection data, which covers more than 10 breathing cycles for most patients. Breath holding is uncomfortable or even impossible for someone such as paediatric or lung cancer patients. The large and complex movement of organs during the data acquisition causes much more serious problems in CBCT than the conventional CT.

Considerable efforts have been made to investigate efficient methods to reduce the motion artefacts in conventional CT and other imaging modalities such as magnetic resonance imaging (MRI) and positron emission tomography (PET) (Atalar and Onural 1991, Buhler *et al* 2004, Crawford *et al* 1996, Cuppen *et al* 1985, Dhanantwari *et al* 2001, Ritchie *et al* 1996, Wang and Vannier 1995, Willis and Bresler 1995). Wang and Vannier (1995) developed a patient motion estimation and compensation technique for helical CT systems, and showed good simulation results, but it was limited to translational motion of the whole patient and did not extend to organ motion. Willis and Bresler (1995) cast the motion artefact problem as a time-varying tomography problem and proposed special-purpose hardware to optimally sample the spatially and temporally band-limited CT signal space. A parametric model for the respiratory motion was first used in MRI, and the motion artefacts were successfully reduced by modifying the reconstruction algorithm (Atalar and Onural 1991, Cuppen *et al* 1985). Crawford *et al* (1996) brought the concept into CT imaging and derived an exact reconstruction formula for motion compensation for CT scans for parallel-beam projections. The method is referred to as the CTX method. In this approach, the respiratory motion was based on magnification and displacement of the object, and the backprojection was performed in a reference frame that moved according to the motion model. Ritchie *et al* (1996) pointed out that the usefulness of CTX was limited by the fact that the time-varying magnification model was not valid for motion in the chest. They extended the method with a physiologically more correct model for respiratory motion and applied the CTX time-varying model on a local basis using *pixel-specific backprojection* (PSBP). A method for estimating the in-plane motion of every pixel in the image at the time each projection is acquired was also developed. However, the method depended heavily on estimating the in-plane motion at individual manually determined node points. Motion correction algorithms that assume a motion model work well when the motion conforms to the model but have limited success when it does not (Ablitt *et al* 2004, Crawford *et al* 1996, Cuppen *et al* 1985, Linney and Gregson 2001). Furthermore, the CTX-based methods cannot be applied to deal with the general 3D motion in CT imaging.

Recently, four-dimensional (4D) CT has gained popularity in guiding radiation treatment in order to explicitly account for the respiratory motion (Low *et al* 2003, Pan *et al* 2004, Rietzel *et al* 2005, Vedam *et al* 2003). With multiple scans at each patient couch position, it generates a series of phase images with respect to the motion data acquired by a real-time positioning system during the scan. The phases can be used to derive a *patient-specific deformation field*,

which accurately models the motion pattern of the patient (Brock *et al* 2003, Schreibmann *et al* 2005). Efforts are being made to use the 4D patient model for time-resolved radiation therapy planning (Keall *et al* 2005, Trofimov *et al* 2005, Webb 2005). In this work, we explore the feasibility of incorporating the 4D patient data into the image reconstruction process to obtain phase-resolved CBCT images. Incorporation of a customized object motion model in CT image reconstruction has been investigated previously (Crawford *et al* 1996, Ritchie *et al* 1996) with the goal of reducing motion artefacts. With the use of the motion information derived from the same patient, the proposed method should lead to a more accurate and robust solution to the problem. In the following, a modified filtered backprojection (FBP) algorithm is developed for the simplest two-dimensional (2D) parallel-beam geometry, and validated with translational and more complex motion, with and without Gaussian noise. A modified Feldkamp algorithm is then implemented for CBCT, and tested with a more realistic deformable phantom constructed from a patient's 4D CT images.

2. Methods and materials

2.1. Reconstruction with deformation field

For 2D parallel geometry, the projection of an object $g(x, y)$, $R(\theta, p)$, at gantry rotational position θ and projection distance p is given by

$$R(\theta, p) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y)\delta(x\cos\theta + y\sin\theta - p)\,\mathrm{d}x\,\mathrm{d}y, \tag{1}$$

where $\delta(.)$ is the Dirac delta-function. The image reconstruction gives a band-limited estimation of the object, $g_B(x, y)$ as follows,

$$g_B(x, y) = \frac{1}{2}\int_0^{2\pi} \int_{-\infty}^{\infty} R(\theta, p) f(x\cos\theta + y\sin\theta - p)\,\mathrm{d}p\,\mathrm{d}\theta, \tag{2}$$

where $f(s) = \int_{-\omega_c}^{\omega_c} |\omega| W(\omega)\exp(2\pi \mathrm{j}\omega s)\,\mathrm{d}\omega$ is the filter function. Different types of filters can be obtained by different window function $W(\omega)$ designs. In practice, equation (2) is usually implemented by the following 'filtered backprojection' steps,

$$\tilde{R}(\theta, p') = \int_{-\infty}^{\infty} R(\theta, p) f(p' - p)\,\mathrm{d}p, \tag{3}$$

$$g_B(x, y) = \frac{1}{2}\int_0^{2\pi} \tilde{R}(\theta, x\cos\theta + y\sin\theta)\,\mathrm{d}\theta, \tag{4}$$

where (3) is the filtered step implemented by convolution (or can be alternatively implemented by Fourier transform), and (4) is the backprojection step.

Now if the object moves during the scan, the projection data acquired at each angle $\theta_i, i = 0, 1, 2, \ldots, N-1$, actually correspond to a series of 'different objects' g_{θ_i} (or more precisely, different status of the object), which can be described as the object at the first phase g_0 (or equivalently, at the first projection angle) being deformed by a time-dependent transformation $\mathbf{T}_{\theta_i}$ (i.e., a 4D motion model), so that

$$g_{\theta_i}(x, y) = \mathbf{T}_{\theta_i}[g_0](x, y). \tag{5}$$

We assume that the transformation $\mathbf{T}_{\theta_i}$ is known for each projection angle, so is the 'inverse' transformation $\mathbf{T}'_{\theta_i}$, $g_0(x, y) = \mathbf{T}'_{\theta_i}[g_{\theta_i}](x, y)$. It will be discussed later on how to use deformable registration to obtain these transformations.

In order to apply the motion model to reconstruction, we rewrite the reconstruction formula (4) in the following discrete form,

$$g_B(x, y) = \frac{\pi}{N} \sum_{i=1}^{N} \tilde{g}_{\theta_i}(x, y), \tag{6}$$

where $\tilde{g}_{\theta_i}(x, y) = \tilde{R}(\theta_i, x \cos\theta_i + y \sin\theta_i)$ is essentially the backprojection of filtered projection at angle θ_i, and can be regarded as an intermediately reconstructed object with a single-angle projection. By summing all these intermediate objects together from all angles, some pixels are enhanced and some pixels are cancelled out, and a reconstructed image can be obtained. However, when motion is present, the corresponding pixels of the intermediate objects are misplaced, and the summation will result in blurring, doubling or other distortions. Similar to the assumption used by Ritchie *et al* (1996), one can assume that local correction was a valid approximation in CT reconstruction and the backprojection can be performed in a deformed reference. Therefore, we propose heuristically to deform the intermediate objects to the same phase before doing the summation:

$$g_B(x, y) \approx \frac{\pi}{N} \sum_{i=1}^{N} \mathbf{T}'_{\theta_i}[\tilde{g}_{\theta_i}(x, y)]. \tag{7}$$

Note that $\tilde{g}_{\theta_i}(x, y)$ of similar phases can be grouped together before the deformation is performed.

2.2. Extension to cone-beam geometry

Extension of the motion compensation method to other geometries such as fan-beam or cone-beam can be done in the same way when backprojection-type algorithms (Feldkamp *et al* 1984, Lange and Carson 1984) are used for reconstruction. In particular for the circular orbit CBCT, the Feldkamp algorithm commonly built in commercial machines is modified to accommodate the motion effects.

Following the notation in Wang *et al* (1993), with $R(\theta, p, \varsigma)$ denoting the cone-beam projection of an object $g(x, y, z)$ at gantry rotational position θ and detector bin (p, ς), the Feldkamp algorithm can be expressed as the 'weighted filtered backprojection',

$$g(x, y, z) = \frac{1}{2} \int_0^{2\pi} \frac{\rho^2}{(\rho - s)^2} \tilde{R}(\theta, u, v)\, \mathrm{d}\theta \tag{8}$$

$$\tilde{R}(\theta, u, v) = \int_{-\infty}^{\infty} \frac{\rho}{\sqrt{\rho^2 + p^2 + v^2}} R(\theta, p, v) f(u - p)\, \mathrm{d}p, \tag{9}$$

where ρ is the source to axis distance (SAD), $s = -x \sin\theta + y \cos\theta$, $\tilde{R}(\theta, u, v)$ is the filtered projection, and (u, v) is the intersecting point on the detection plane of the ray coming from the cone vertex through the reconstruction point (x, y, z), and $f(.)$ is the filter function. More details can be found in the generalized Feldkamp algorithm of Wang *et al* (1993). Again, equation (8) can be rewritten in a similar form to (6), and we can apply the deformation to the single-angle backprojected objects before the summation operation and obtain the following approximate formula:

$$\tilde{g}_{\theta_i}(x, y, z) = \frac{\rho^2}{(\rho - s)^2} \tilde{R}(\theta_i, u, v), \tag{10}$$

$$g(x, y, z) \approx \frac{\pi}{N} \sum_{i=1}^{N} \mathbf{T}'_{\theta_i}[\tilde{g}_{\theta_i}(x, y, z)]. \tag{11}$$

2.3. Motion model derivation

In radiotherapy clinics, the CBCT on-board imager aims to accurately position the patient in the treatment room. For a tumour under the influence of respiration motion, 4D CT has been adopted to extract organ motion information for 4D treatment planning (Keall *et al* 2005, Trofimov *et al* 2005, Webb 2005). The 4D patient model derived from the 4D CT data can be applied to correct motion artefacts in CBCT imaging.

A deformable image registration model is required to obtain the 4D patient model from the 4D CT data. For convenience, we adopted the free-form Spline model (BSpline) (Mattes *et al* 2003) in this work to register the 4D CT images. The simplicity and yet accuracy of the BSpline method make it a preferred tool for many clinical applications (Lian *et al* 2004, Rohlfing *et al* 2003, Rueckert *et al* 1999). In this model, a lattice of user-defined nodes is overlaid on the image. Each node contains a deformation vector, whose components are determined by an optimization procedure. The deformation at any point of the image is calculated by Spline interpolation of the adjacent node values. One advantage of the BSpline model is that the nodes are *locally* controlled, such that the displacement of an interpolation point is influenced only by the nearest grid points and changing a lattice node only affects the transformation regionally, making it efficient in describing local deformations. Suitable node deformations are obtained using the gradient-based L-BFGS algorithm (Liu and Nocedal 1989, Schreibmann and Xing 2005), which iteratively varies the deformations until the registration metric, a mathematical measure of similarity between images, is minimized. The normalized cross correlation metric was used as the metric in this study. With deformable registration, all phases can be registered to one particular phase, for example at $t = 0$, resulting in a series of transformations representing a temporal sequence of 3D deformation fields, or in other words, a 4D model of organ motion.

2.4. Computer simulations

The performance of the proposed algorithm was tested by computer simulations in this work for both in-plane motion and 3D motion. First, the in-plane motion was studied with a 2D digital dynamic phantom, capable of both rigid translational motion and non-rigid shape change. As shown in figure 1, the phantom consisted of several ellipses of different sizes and densities. The whole phantom varied its shape constantly during the data acquisition, and an internal circle was also moving vertically with an amplitude of 1.5 cm. The motion period was 4.0 s, of which the inhale stage took 3.0 s and the exhale stage took 1.0 s. The phantom size was 512×512 pixels, with the pixel size of 1 mm. The projections were simulated by line integrals of the phantom density with parallel-beam geometry. For the slow-rotating CT, the simulated 360° scan took 40.0 s and generated 512 equally time-spaced projections. In other words, the projections were acquired at time $t = i \times 40/511\,\text{s}, i = 0, 1, \ldots, 511$, with the phantom moving correspondingly. The time it took for acquiring each single projection was neglected. For comparison, a simulated full scan of the stationary phantom characterizing its state at $t = 0$ was also performed.

Furthermore, Gaussian noise was added into the imaging process to test the robustness of the proposed motion correction strategy. Specifically, 30% Gaussian noise was added into each phase of the motion phantom as well as the simulated projections. Therefore, the deformation fields (or motion model) derived from the phantoms were subjected to the impact of the noise, and the errors of the derived motion model will subsequently propagate into the image reconstruction process.

Finally, 3D circular orbit CBCT was simulated with a more realistic digital phantom to test the motion correction method. In this study, the phantom was constructed with 4D CT images

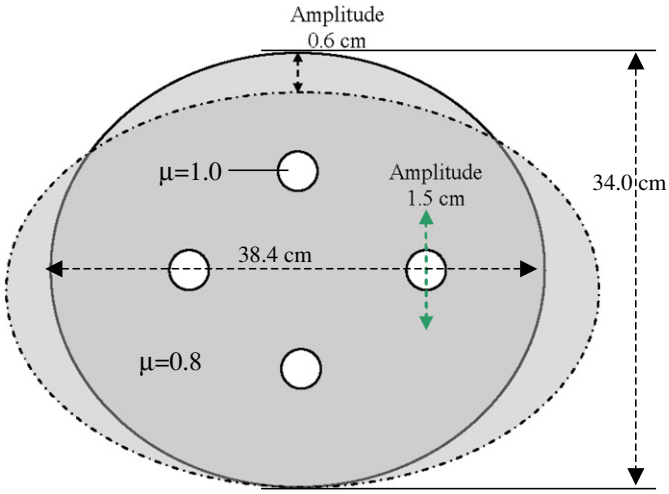


Figure 1. A 2D dynamic phantom for CT simulations with a complex mode. The whole phantom changes its shape as indicated by the dotted line with a period of 3.52 s. The volume of the phantom was kept constant during the movement. The inner right circle moves vertically with the same period (plot was not drawn to a real scale).

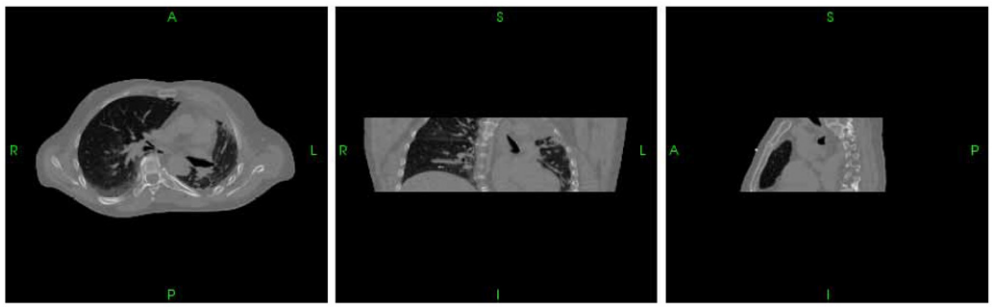


Figure 2. A 3D deformable phantom for CBCT simulations. The phantom was constructed from 4D CT images of a lung patient. The CT numbers were converted to the attenuation coefficient and the top row is the volume rendering of the data. From left to right it contains the transaxial, coronal and sagittal images of the phantom.

of a lung patient. The CT numbers (Hounsfield unit) were converted into the attenuation coefficient and a set of eight volumetric data corresponding to eight respiratory phases of the patient were obtained, each with an array size of 256 × 256 × 64 voxels. Figure 2 shows the phantom at the first phase with three views and with volume rendering. In the simulations, the cone-beam source to detector distance was 1500 mm, and 1000 mm to the centre of rotation. The detector size was 512 mm × 128 mm. A total of 256 projections were simulated, with the phantom moving from one phase to another at subsequent projection angles. For comparison, a static CBCT simulation was also performed with the same phantom for the first phase. The free-form BSpline deformable registration was applied to the eight phases of the phantom to extract the motion model, and the projection data were then reconstructed with the conventional and modified Feldkamp algorithm.

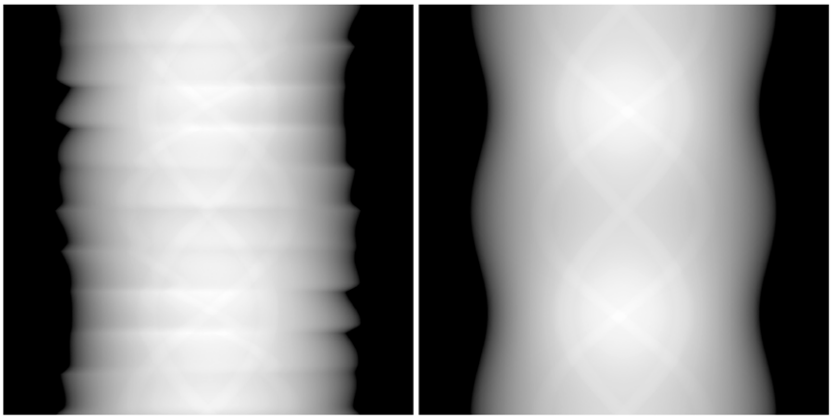


Figure 3. Comparison of the sinogram with moving phantom and stationary phantom. Left: simulated projections with the 2D dynamic phantom; right: simulated projections with the same phantom with the motion 'switched off'.

3. Results

3.1. 2D simulations

Figure 3 shows the simulated motion corrupted projections with the 2D dynamic phantom, as well as the projections of the same phantom with motion 'switched off'. The inconsistency in the sinogram of the moving phantom is clearly seen in the picture. To compensate for the motion in the image reconstruction, we first derived the motion model for the dynamic phantom. The deformation field at each phase with respect to the first phase was obtained by the BSpline deformable registration described earlier. As an example, we illustrate in figure 4 the deformation fields of phase $i = 10$ relative to phase $i = 0$ at each region of the phantom. The arrows in the figure show the direction of the movement at each pixel and the colour indicates the amplitude of the deformation vectors.

By applying the deformation in the backprojection step of the reconstruction, the images with reduced motion artefacts can be obtained. The reconstructed images with and without motion correction are shown in figure 5, in which the left one is the phantom, the middle is the reconstructed image without motion correction, and the right one is the reconstructed image with motion correction. It is observed that the outer boundary of the motion phantom is corrected with the approach, and the shape of the small moving circle is also restored. Figure 6 shows the vertical profiles through the moving circle, in which it is found that the intensity of the moving circle is in accordance with the phantom value.

In order to test the robustness of the proposed method, we further added 30% Gaussian noise into the simulation to create uncertainties in both projections and the derived deformation fields. Although the deformable registration was affected by the noise (see figures 7 and 4), it is found that the density of the moving part of the phantom was recovered correctly and the outer boundary of the phantom was corrected as well, see figure 8.

3.2. 3D cone-beam simulations

The proposed motion compensation method was further tested under cone-beam geometry with a 4D anthropomorphic digital phantom. Figure 9 shows the eight phases of the phantom. The

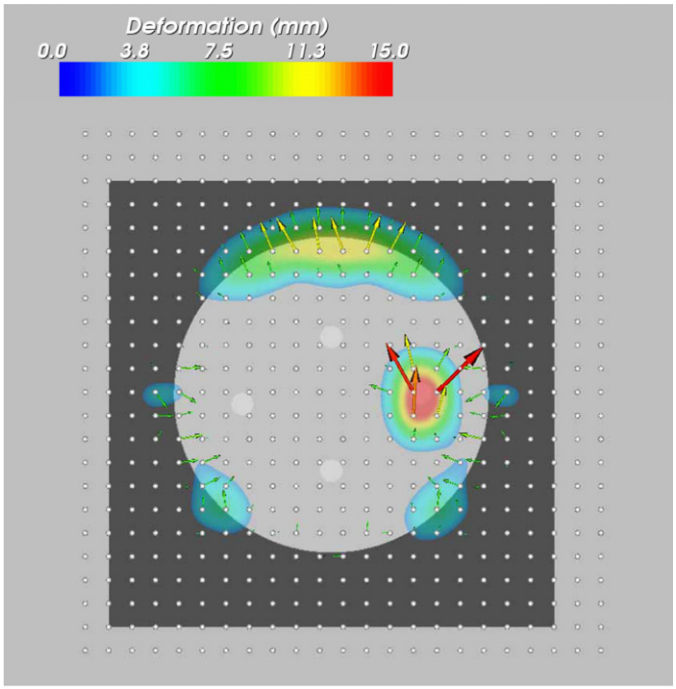


Figure 4. Deformation field obtained by BSpline registration of phase $i = 10$ and phase $i = 0$ for the motion phantom shown in figure 1.

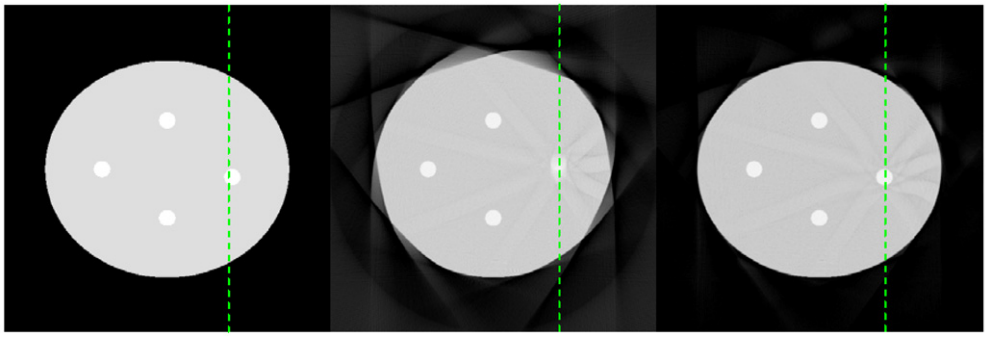


Figure 5. Phantom and reconstructed images. Left: phantom image; middle: reconstructed image without correction; right: reconstructed image with motion correction.

derived deformation field for phase 1 to phase 0 is illustrated in figure 10. The reconstructed images with the conventional and modified Feldkamp algorithm are presented in figure 11, in which the first column shows the reconstructed images of CBCT projection of the moving phantom with the conventional Feldkamp algorithm, and the middle column shows the same data reconstructed with the proposed motion compensation algorithm. For comparison, the static simulation with the phantom at phase 0 was reconstructed with the conventional Feldkamp algorithm and is shown in the last column. From top to bottom in the figure are the axial, coronal and sagittal views, and the last row shows zoom-in images of the region of

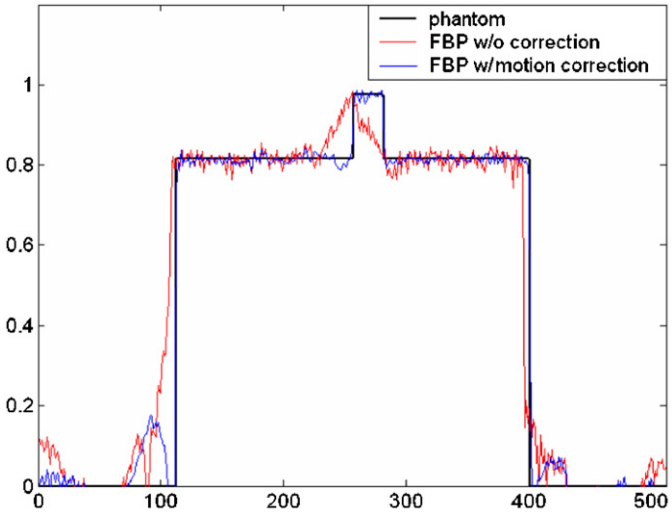


Figure 6. Vertical profiles through the inner moving circle for images shown in figure 5.

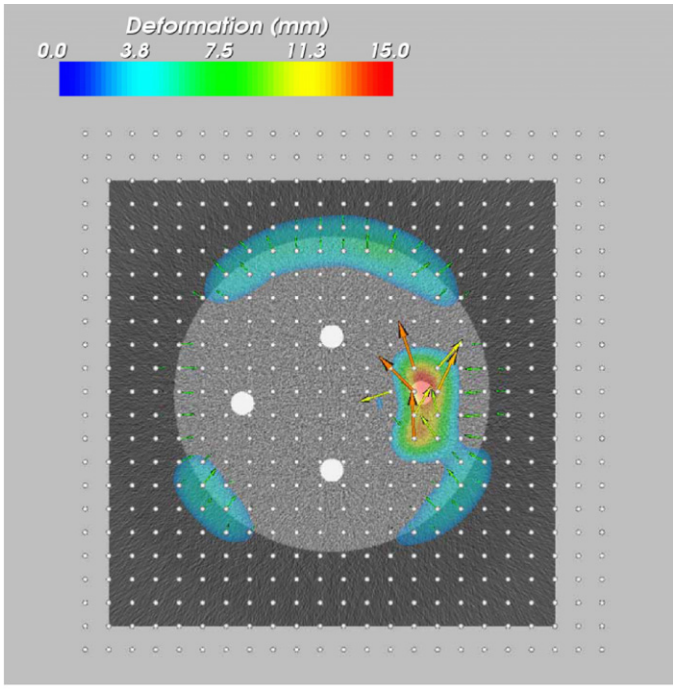


Figure 7. Deformation field obtained in the presence of Gaussian noise by BSpline registration for phase $i = 10$ and phase $i = 0$ for the 2D dynamic phantom.

interest illustrated in the axial view. It is found that the motion artefacts were reduced and the tumour intensity and shape were restored by our proposed method.

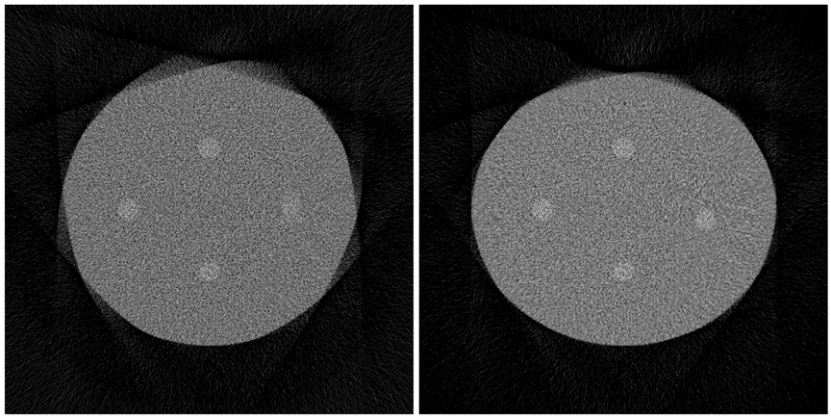


Figure 8. Reconstructed images with noise present. Left: reconstructed image without correction; right: reconstructed image with motion correction. The derived motion model was affected by the noise; however, the reduction of the tumour distortion was still observed.

4. Discussion and conclusion

There are generally two approaches to the removal of organ motion artefacts in CT imaging. The first is to sort the projections according to their phases and then reconstruct the sorted projections separately to obtain time-resolved CT images. Another approach is to incorporate the patient's motion model, which describes how each point of the object moves during the data acquisition, into the image reconstruction process. For CBCT, the first approach is less adequate because there are not sufficient projects for a given phase to warrant high quality 4D CBCT images, unless multiple gantry rotations are performed. In this work, we have demonstrated the feasibility of incorporating a 4D motion model into the CBCT image reconstruction. The motion model used in this work is derived on a patient-specific basis, therefore could be more suitable than other mathematical models for a patient study. In radiotherapy clinics, 4D CT has been adopted to identify the tumour motion and to guide the 4D treatment planning process. Using the 4D CT data for the removal of motion artefacts in CBCT acquisition seems to be logical and fits well with the data flow of modern image-guide radiotherapy (IGRT).

Incorporating a motion model into the de-convolution process is a general approach in estimation theory and has been applied to CT and MRI reconstruction and radiation dose optimization (Li and Xing 2000, Pugachev and Xing 2002). In this process, the motion model serves as *a priori* knowledge of the system and provides valuable partial guidance to the searching process. It should be noted that a potential shortcoming of CBCT reconstruction with inclusion of the 4D patient model derived at an earlier time point is that it does not consider possible change in the patient's organ motion pattern. Indeed, the motion pattern may change in both spatial and temporal domains to some degree from the planning CT when the treatment CBCT is acquired. This represents an important challenge not only to the 4D CBCT reconstruction proposed here, but also to the implementation of 4D radiation therapy (e.g., tumour tracking based on 4D planning). Other uncertainties of the motion model include spatial and temporal errors, for example, due to accuracy in aligning the CBCT coordinate system with the 4D CT coordinate system and in determining the phase of each individual CBCT projection or 4D CT images. However, unless the organ motion pattern is completely reversed, we anticipate that the proposed technique will help to reduce the motion artefacts

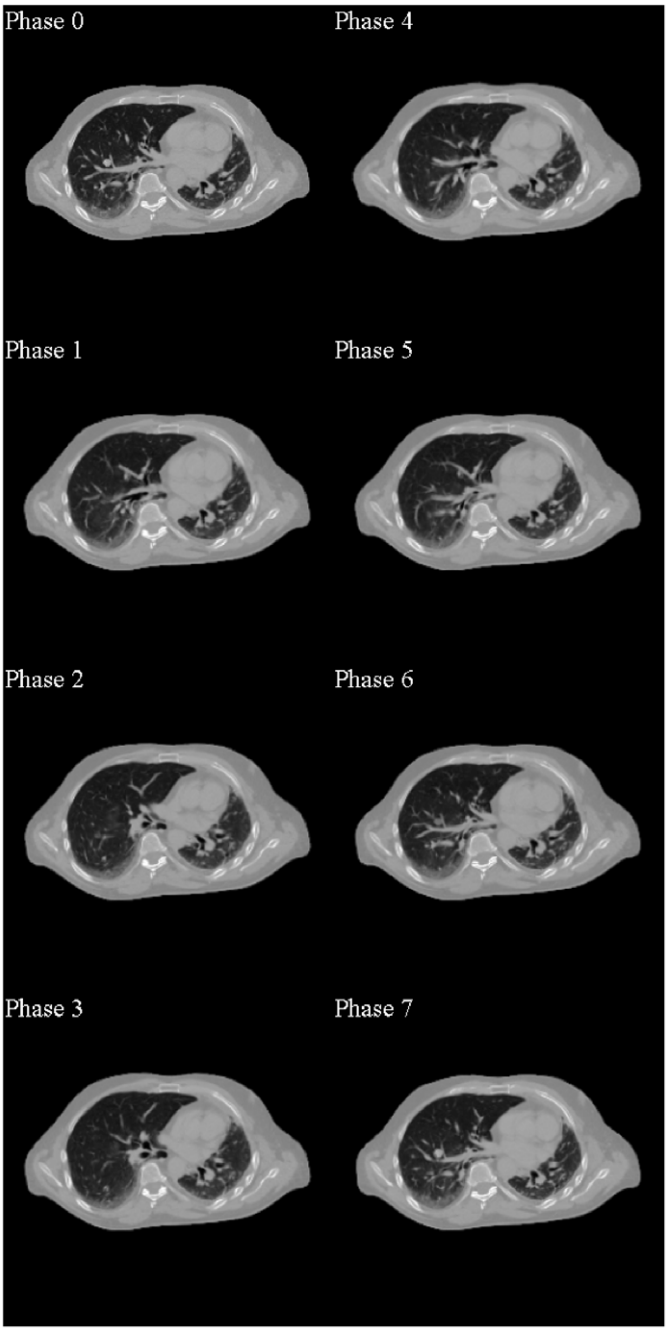


Figure 9. The eight phases of the 4D anthropomorphic phantom.

even if a slight change in the motion pattern occurs. In practice, a consistent respiratory pattern should be maintained through, for example, a proper voice coaching, during the CBCT data acquisition process. This research also highlights the importance of the development of

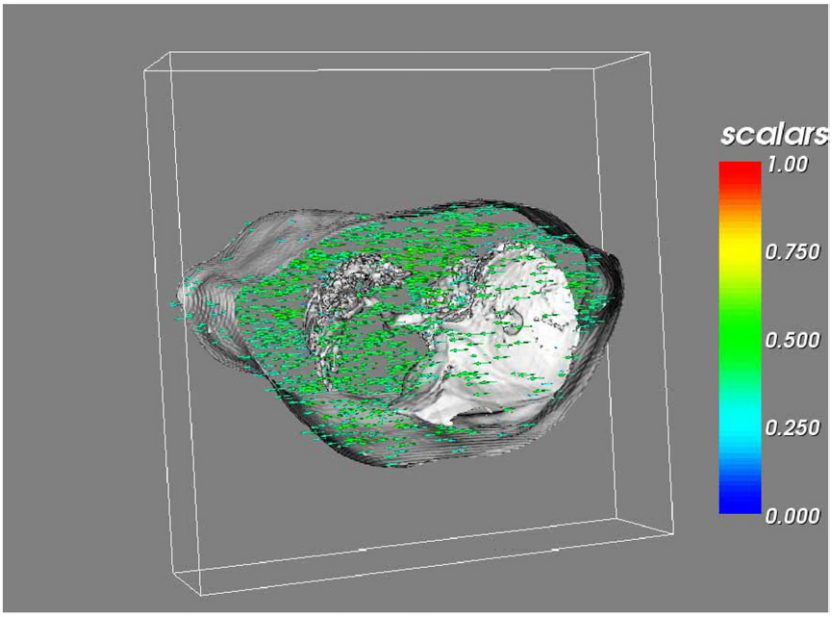


Figure 10. The derived deformation fields of the 3D phantom at phase 1 relative to phase 0 by BSpline deformable registration.

a robust organ motion tracking system in the future. With this, one may obtain the 4D patient model during the 4D CBCT process and then incorporate it into the CBCT image reconstruction process. A study to systematically examine the sensitivity of the 4D CBCT reconstruction against various possible variations of the patient motion model should be of scientific and practical interest. Because of the large scope of the work, we focus this manuscript on setting up the theoretical framework of 4D CBCT reconstruction and leave the sensitivity study for the future investigations.

In the proposed approach, when deformation field is applied with backprojection, the reconstructed image is locally corrected, which is same as Rietch *et al* did with the pixel-specific 'magnification and displacement' motion model. It should be noted that the local correction does not correct the artefacts that are left to other regions, for example, regions that have no motion at all but still suffer from the artefacts produced by the moving part. In general, an iterative method could be utilized to incorporate the motion model for image reconstruction with better motion artefacts correction. The ordered subset convex (OSC) algorithm (Kole and Beekman 2005, Manglos *et al* 1992) could be a good choice for the CBCT task, since the subset can be naturally correlated to the phase set for an efficient implementation. However, the presented approach should be much faster than iterative methods with sufficient restoration of the moving target information.

In summary, we have proposed a motion correction method for reconstructing CT images under the influence of intra-fraction organ motion. The technique is conceptually interesting and may find a natural application in routine CBCT-based patient positioning for improved accuracy. The computation is efficient, involving the standard filtered backprojection step and deformation for each projection phase. The method deals with not only in-plane motion, but also 3D complex motion via deformable registration. It enhances the quality of the

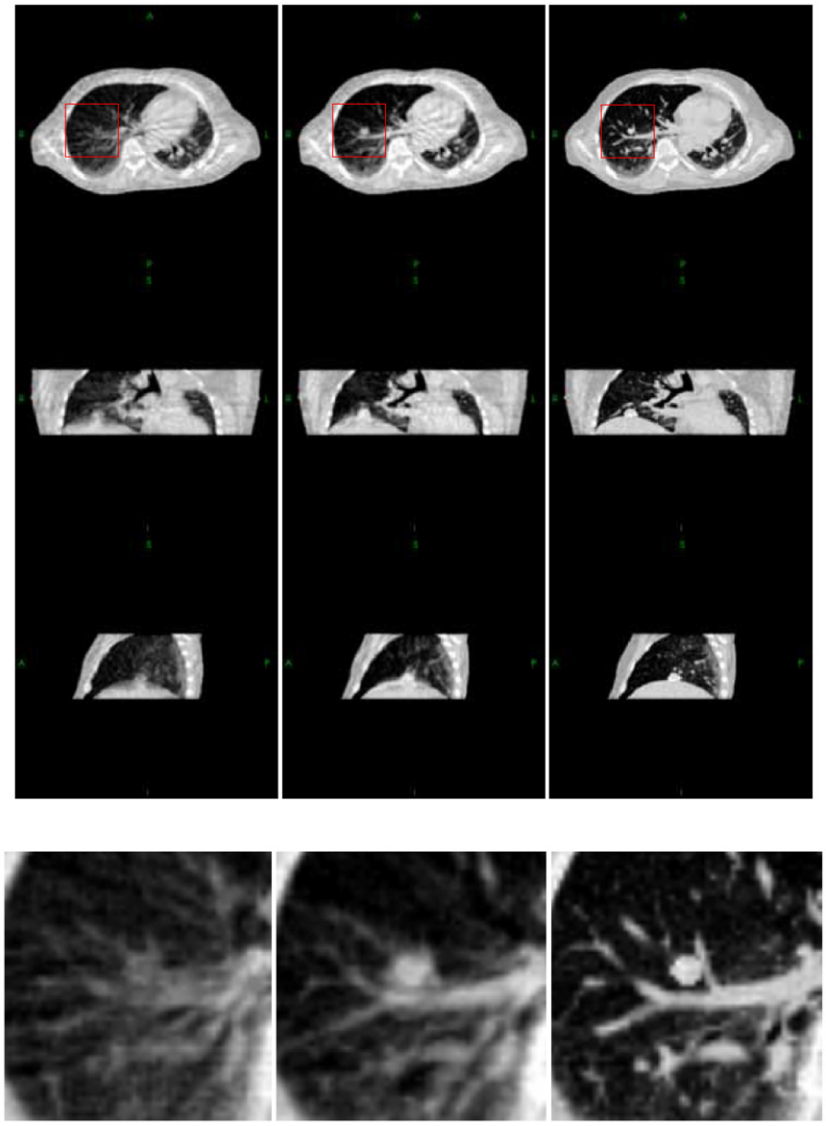


Figure 11. Reconstructed images for the CBCT simulations. Left column: reconstructed images without motion correction; middle column: images with motion correction; right column: images reconstructed with the conventional Feldkamp algorithm for projections with stationary phantom (the 3D phantom at phase 0). From top to bottom are the axial, coronal and sagittal views, and the last row contains the zoom-in images of the region of interest shown in the first row.

reconstructed image, corrects for the density, and restores the shape of the moving tumour, therefore, can provide better target localization.

Acknowledgments

This research is supported in part by Department of Defense (BC021705) and the National Cancer Institute (1 R01 CA98523-01).

References

Ablitt N A, Gao J, Keegan J, Stegger L, Firmin D N and Yang G Z 2004 Predictive cardiac motion modeling and correction with partial least squares regression *IEEE Trans. Med. Imaging* **23** 1315–24

Atalar E and Onural L 1991 A respiratory motion artifact reduction method in magnetic resonance imaging of the chest *IEEE Trans. Med. Imaging* **10** 11–24

Brock K K, McShan D L, Ten Haken R K, Hollister S J, Dawson L A and Balter J M 2003 Inclusion of organ deformation in dose calculations *Med. Phys.* **30** 290–5

Buhler P, Just U, Will E, Kotzerke J and Hoff J V D 2004 An accurate method for correction of head movement in pet *IEEE Trans. Med. Imaging* **23** 1176–85

Crawford C R, King K F, Ritchie C J and Godwin J D 1996 Respiratory compensation in projection imaging using a magnification and displacement model *IEEE Trans. Med. Imaging* **15** 327–32

Cuppen J J M, Groen J P, Kleef J J E I D and Tuithof H A 1985 Reduction of motion artifacts by data processing *Proc. Soc. Magn. Reson. Med.* 962–3

Dhanantwari A C, Stergiopoulos S and Iakovidis I 2001 Correcting organ motion artifacts in x-ray ct medical imaging systems by adaptive processing. I. Theory *Med. Phys.* **28** 1562–76

Feldkamp L A, Davis L C and Kress J W 1984 Practical cone-beam alogirthm *J. Opt. Soc. Am.* **1 A** 612–9

Jaffray D A, Drake D G, Moreau M, Martinez A A and Wong J W 1999 A radiographic and tomographic imaging system integrated into a medical linear accelerator for localization of bone and soft-tissue targets *Int. J. Radiat. Oncol. Biol. Phys.* **45** 773–89

Jin J Y and Yin F F 2005 Time delay measurement for linac based treatment delivery in synchronized respiratory gating radiotherapy *Med. Phys.* **32** 1293–6

Keall P J, Joshi S, Vedam S S, Siebers J V, Kini V R and Mohan R 2005 Four-dimensional radiotherapy planning for dmlc-based respiratory motion tracking *Med. Phys.* **32** 942–51

Kole J S and Beekman F J 2005 Parallel statistical image reconstruction for cone-beam x-ray ct on a shared memory computation platform *Phys. Med. Biol.* **50** 1265–72

Lange K and Carson R 1984 Em reconstruction algorithms for emission and transmission tomography *J. Comput. Assist. Tomogr.* **8** 306–16

Li J G and Xing L 2000 Inverse planning incorporating random organ motion *Med. Phys.* **27** 1368–72

Lian J, Xing L, Hunjan S, Dumoulin C, Kim D, Spielman D and Daniel B 2004 Mapping of the prostate in endorectal coil-based mri/mrsi and ct: A deformable registration and validation study *Med. Phys.* **31** 3087–94

Linney N C and Gregson P H 2001 Organ motion detection in ct images using opposite rays in fan-beam projection systems *IEEE Trans. Med. Imaging* **20** 1109–22

Liu D C and Nocedal J 1989 On the limited memory bfgs method for large scale optimization *Math. Program.* **45** 503–28

Low D A *et al* 2003 A method for the reconstruction of four-dimensional synchronized ct scans acquired during free breathing *Med. Phys.* **30** 1254–63

Manglos S H, Bassano D A, Thomas F D and Grossman Z D 1992 Imaging of the human torso using cone-beam transmission ct implemented on a rotating gamma camera *J. Nucl. Med.* **33** 150–6

Mattes D, Haynor D R, Vesselle H, Lewellen T K and Eubank W 2003 Pet-ct image registration in the chest using free-form deformations *IEEE Trans. Med. Imaging* **22** 120–8

Moseley D J *et al* 2004 Comparison of implanted fiducial markers and cone-beam computed tomography for on-line image-guided radiotherapy of the prostate *Int. J. Radiat. Oncol. Biol. Phys.* **60** S330–S331

Pan T, Lee T Y, Rietzel E and Chen G T 2004 4d-ct imaging of a volume influenced by respiratory motion on multi-slice ct *Med. Phys.* **31** 333–40

Pugachev A and Xing L 2002 Incorporating prior knowledge into beam orientation optimization *Int. J. Radiat. Oncol. Biol. Phys.* **54** 1565–74

Rietzel E, Pan T and Chen G T 2005 Four-dimensional computed tomography: Image formation and clinical protocol *Med. Phys.* **32** 874–89

Ritchie C J, Crawford C R, Godwin J D, King K F and Kim Y 1996 Correction of computed tomography motion artifacts using pixel-specific back-projection *IEEE Trans. Med. Imaging* **15** 333–42

Rohlfing T, Maurer C R Jr, Bluemke D A and Jacobs M A 2003 Volume-preserving nonrigid registration of mr breast images using free-form deformation with an incompressibility constraint *IEEE Trans. Med. Imaging* **22** 730–41

Rueckert D, Sonoda L I, Hayes C, Hill D L, Leach M O and Hawkes D J 1999 Nonrigid registration using free-form deformations: application to breast MR images *IEEE Trans. Med. Imaging* **18** 712–21

Schreibmann E, Chen G T Y and Xing L 2005 Image interpolation in 4D CT using a BSpline deformable registration model *Int. J. Radiat. Oncol. Biol. Phys.* at press

Schreibmann E and Xing L 2005 Narrow band deformable registration of prostate magnetic resonance imaging, magnetic resonance spectroscopic imaging, and computed tomography studies *Int. J. Radiat. Oncol. Biol. Phys.* **62** 595–605

Sidhu K *et al* 2003 Optimization of conformal thoracic radiotherapy using cone-beam ct imaging for treatment verification *Int. J. Radiat. Oncol. Biol. Phys.* **55** 757–67

Sonke J J, Zijp L, Remeijer P and van Herk M 2005 Respiratory correlated cone beam ct *Med. Phys.* **32** 1176–86

Trofimov A, Rietzel E, Lu H M, Martin B, Jiang S, Chen G T and Bortfeld T 2005 Temporo-spatial imrt optimization: Concepts, implementation and initial results *Phys. Med. Biol.* **50** 2779–98

van Herk M J D *et al* 2004 First clinical experience with cone-beam ct guided radiation therapy: Evaluation of dose and geometric accuracy *Int. J. Radiat. Oncol. Biol. Phys.* **60** S196

Vedam S S, Keall P J, Kini V R, Mostafavi H, Shukla H P and Mohan R 2003 Acquiring a four-dimensional computed tomography dataset using an external respiratory signal *Phys. Med. Biol.* **48** 45–62

Wang G, Lin T-H, Cheng P-C and Shinozaki D M 1993 A general cone-beam reconstruction algorithm *IEEE Trans. Med. Imaging* **12** 486–96

Wang G and Vannier M W 1995 Preliminary study on helical ct algorithms for patient motion estimation and compensation *IEEE Trans. Med. Imaging* **14** 205–11

Webb S 2005 The effect on imrt conformality of elastic tissue movement and a practical suggestion for movement compensation via the modified dynamic multileaf collimator technique *Phys. Med. Biol.* **50** 1163–90

Willis N P and Bresler Y 1995 Optimal scan for time-varying tomography i:Theoretical analysis and fundamental limitations *IEEE Trans. Imaging Proc.* 642–53

INSTITUTE OF PHYSICS PUBLISHING PHYSICS IN MEDICINE AND BIOLOGY

Phys. Med. Biol. **50** (2005) 1469–1482 doi:10.1088/0031-9155/50/7/010

Quantitation of the *a priori* dosimetric capabilities of spatial points in inverse planning and its significant implication in defining IMRT solution space*

Z Shou[1], Y Yang[1], C Cotrutz[1], D Levy[2] and Lei Xing[1]

[1] Department of Radiation Oncology, Stanford University, Stanford, CA 94305-5847, USA
[2] Department of Mathematics, Stanford University, Stanford, CA 94305-2125, USA

E-mail: lei@reyes.stanford.edu

Received 4 January 2005, in final form 2 February 2005
Published 16 March 2005
Online at stacks.iop.org/PMB/50/1469

Abstract

In inverse planning, the likelihood for the points in a target or sensitive structure to meet their dosimetric goals is generally heterogeneous and represents the *a priori* knowledge of the system once the patient and beam configuration are chosen. Because of this intrinsic heterogeneity, in some extreme cases, a region in a target may never meet the prescribed dose without seriously deteriorating the doses in other areas. Conversely, the prescription in a region may be easily met without violating the tolerance of any sensitive structure. In this work, we introduce the concept of dosimetric capability to quantify the *a priori* information and develop a strategy to integrate the data into the inverse planning process. An iterative algorithm is implemented to numerically compute the capability distribution on a case specific basis. A method of incorporating the capability data into inverse planning is developed by heuristically modulating the importance of the individual voxels according to the *a priori* capability distribution. The formalism is applied to a few specific examples to illustrate the technical details of the new inverse planning technique. Our study indicates that the dosimetric capability is a useful concept to better understand the complex inverse planning problem and an effective use of the information allows us to construct a clinically more meaningful objective function to improve IMRT dose optimization techniques.

(Some figures in this article are in colour only in the electronic version)

* Part of this work was presented in the 14th International Conference on the Use of Computers in Radiation Therapy, Seoul, Korea, 2004.

0031-9155/05/071469+14$30.00

1. Introduction

One of the implicit assumptions in current inverse planning is that all points within a target or sensitive structure are equivalent (Brahme *et al* 1982, Bortfeld *et al* 1990, Cho *et al* 1998, Gopal and Starkschall 2001, Holmes and Mackie 1994, Langer and Leong 1987, Spirou and Chui 1998, Webb 1991, Xing and Chen 1996, Zagars *et al* 2002). In reality, not all voxels have the same chance of complying with the prescription because the dose-limiting factors imposed by the involved sensitive structures are not uniformly distributed in space and the dose delivery is depth dependent. For a given patient and beam configuration, it is practically useful to know the likelihood for each individual point within a target or sensitive structure to meet its dosimetric goal. By understanding this intrinsic property of the system, one can better model the therapeutic plan optimization problem and improve the inverse planning techniques.

In this work, we introduce the concept of dosimetric capability for an arbitrary voxel in a target or sensitive structure to quantify the likelihood for the voxel to meet the specified dose. The capability calculation finds the potentially problematic regions and, more importantly, the degree of problems for these regions to meet their goals, and permits us to purposely modify the penalty strategy during the construction of objective function to minimize the problem. Mathematically, this process is realized by heuristically modulating the importance of the individual voxels according to the *a priori* capability distribution. The inverse planning formalism with dosimetric capability-modulated importance factors is applied to a few specific examples to illustrate the technical details. The approach sheds useful insight into the inverse planning problem and allows us to search for IMRT solutions that would otherwise be inaccessible.

2. Methods and materials

2.1. Definition of dosimetric capability

We quantify the likelihood for a voxel to meet its dosimetric goal by introducing the concept of dosimetric capability. Let us first consider the case of one incident beam. For a target voxel, the capability is assessed by the degree to which the voxel meets the prescription without violating the tolerance of the sensitive structure. The maximum achievable dose, $D^{\mathrm{ach}}(n_\sigma)$, at the voxel n_σ in the target is determined by scaling the intensity of the contributing beamlet to the highest value set by the tolerance of the sensitive structure. Mathematically, the 'capability', η, is defined as

$$\eta(n_\sigma) = \frac{D^{\mathrm{ach}}(n_\sigma)}{D_\sigma^{\mathrm{pre}}}. \tag{1}$$

The evaluation of equation (1) is straightforward for the case of a single incident beam or when there is no overlap of beamlets at the dose-limiting voxel in the sensitive structure. For multi-field IMRT, the $D^{\mathrm{ach}}(n_\sigma)$ is determined not only by those beamlets that directly intercept the voxel n, but also possibly by other beamlets irradiating different parts of the target. The coupled system can be described by a set of linear equations (which can be easily written from the above definition of the maximum achievable dose and the dose as a function of beamlet weights) and the $D^{\mathrm{ach}}(n_\sigma)$ of a target voxel will be obtained by optimizing the linear system, under the condition that the dose at any sensitive structure voxel is equal to its tolerance. The beam profiles so obtained deliver the maximum dose to the target without violating the tolerances of the sensitive structures. Mathematically, the system equations are underdetermined and a Cimmino algorithm (Stark and Yang 1988, Xiao *et al* 2000), which

was first applied to radiation therapy by Starkschal and Eifel (1992), can be used to find the solution. After assigning all beamlets with high enough initial intensities (say, two or three times the intensity values that deliver the prescribed dose to the target) and dose calculation, the calculation consists of the following steps.

(a) Choose a beamlet in a beam and locate the voxels in sensitive structures that are traversed by the beamlet.
(b) For each located voxel in the sensitive structures, check if the tolerance is exceeded. If yes, decrease the value of the beamlet based on

$$w_{jm}^{k+1} = w_{jm}^{k} + \lambda \sum_{n_\sigma} c_{n_\sigma}(w^k) d_{jm}(n_\sigma), \tag{2}$$

with

$$c_{n_\sigma}(w^k) = \begin{cases} \dfrac{\left(D_\sigma^{\mathrm{pre}} - D_c^k(n_\sigma)\right)}{\sum_{jm} d_{jm}^2(n_\sigma)} & \text{if } D_c^k(n_\sigma) > D_\sigma^{\mathrm{tol}} \text{ and } n_\sigma \in \text{sensitive structures} \\ 0 & \text{otherwise.} \end{cases}$$

(c) Update the dose.
(d) Repeat steps (b) and (c) until all voxels intercepted by the beamlet are checked.
(e) Repeat (a)–(d) for the next beamlet.

The above calculation is repeated until the doses in sensitive structures are equal to their tolerances. The relaxation parameter λ is generally set to a small value ($0.02 < \lambda < 0.1$) to ensure a smooth updating of the beamlet weights. The beamlet-by-beamlet updating scheme is similar to the simultaneous iterative inverse planning technique (SIITP) (Xing and Chen 1996). However, the goal here is to search for the beam profiles that deliver the highest achievable doses in the target, $D^{\mathrm{ach}}(n_\sigma)$, without violating the dose tolerance of the involved sensitive structures (in other words, any increase in the beamlet weights would lead to a dose exceeding the tolerance in one or multiple points inside a sensitive structure). Mathematically, the above calculation is an iterative projection algorithm converging to a feasible solution if the constraints are satisfied, otherwise to a compromise solution that minimizes the weighted sum of squares of deviations from the tolerance doses if dose constraints are violated.

The dosimetric capability of a voxel in a sensitive structure is characterized by the minimum achievable dose at the voxel needed in order to administer the prescribed dose to the target voxels. For consistency, we denote the minimum achievable dose by $D^{\mathrm{ach}}(n_\sigma)$ (where n_σ represents a voxel in the sensitive structure) and quantify the dosimetric capability of the voxel according to

$$\eta(n_\sigma) = \frac{D_\sigma^{\mathrm{tol}}}{D^{\mathrm{ach}}(n_\sigma)}. \tag{3}$$

For a voxel in a sensitive structure, the lower the minimum achievable dose, the more 'capable' the voxel is. The evaluation of equation (3) is once again straightforward for the case with a single incident beam. To obtain the minimum achievable dose, we first set the corresponding beamlet to such an intensity that the prescribed dose is delivered to the target voxels, and then evaluate the dose, $D^{\mathrm{ach}}(n_\sigma)$, at n_σ in the sensitive structure. For the case of multiple incident beams, a set of linear equations needs to be solved under the condition that the target voxels receive their prescription doses. Similar to the target case, the system equations become undetermined when there are multiple incident beams and a Cimmino algorithm is used to find the beam profiles that yield the lowest achievable dose in the sensitive structures. We first assign all beamlets with low enough initial intensities so that doses delivered to the target voxels are less than the prescription, and then do the following.

(a) Choose a beamlet in a beam and locate the voxels in the target that are traversed by the beamlet.

(b) For each located voxel in the target, check if the dose prescription is exceeded. If yes, lower the beamlet weight based on equation (2) with

$$c_{n_\sigma}(w^k) = \begin{cases} \dfrac{\left(D_\sigma^{\text{pre}} - D_c^k(n_\sigma)\right)}{\sum_{jm} d_{jm}^2(n_\sigma)} & \text{if } D_c^k(n_\sigma) < D_\sigma^{\text{pre}} \text{ and } n_\sigma \in \text{target} \\ 0 & \text{otherwise.} \end{cases}$$

(c) Update the doses.

(d) Repeat steps (b) and (d) until all voxels intercepted by the beamlet are checked.

(e) Repeat (a)–(d) for the next beamlet.

The above calculation proceeds iteratively until the target dose prescription is completely met for all voxels. We emphasize that the beam profiles obtained above are not intended to approximate the optimal fluence profiles for IMRT treatment. Instead, they are obtained purely for the purpose of evaluating the dosimetric capabilities of voxels in a target or a sensitive structure. Once the capability maps are obtained, they can be incorporated into an inverse planning procedure.

2.2. Incorporating dosimetric capability into inverse planning

The capability distribution contains information about the intrinsic heterogeneity of the dosimetric capability and reveals which points are likely to violate the prescription. The information provides a guiding map for the inverse planning algorithm to differentially deal with the regions having different chances of meeting their dosimetric goals. Our strategy is to assign a higher penalty (or high local importance) to those voxels with lower capabilities (those voxels are likely to have a dose lower than the prescription if they are in the target volume, or higher than the tolerance if they are located in a sensitive structure). While it is not difficult to intuitively conceive the general behaviour of the relation between local importance and the capability, the specific form of the relation is a matter of experimenting. The relation between the local importance and the capability used in our study is

$$\tilde{r}(n_\sigma) = (1 + g(n_\sigma)^2) r_\sigma, \tag{4}$$

where $g(n_\sigma)$ is empirically defined as

$$g(n_\sigma) = \begin{cases} \dfrac{3}{\left(\eta_\sigma^{\max} - 1\right)}(\eta(n_\sigma) - 1) & n_\sigma \in \text{target} \quad \text{and} \quad \eta(n_\sigma) \geqslant 1 \\ \dfrac{3}{\left(\eta_\sigma^{\min} - 1\right)}(\eta(n_\sigma) - 1) & n_\sigma \in \text{target or sensitive structure} \quad \text{and} \quad \eta(n_\sigma) < 1. \end{cases} \tag{5}$$

The capability map and the corresponding $\eta_\sigma^{\max}$ or $\eta_\sigma^{\min}$ are obtained for each structure. For a sensitive structure, if $\eta(n_\sigma) > 1$, it means that the voxel has no difficulty meeting its dosimetric goal and $g(n_\sigma)$ is set to zero. The differential penalty scheme allows the system to suppress potential hot spots in the sensitive structures and boost the potential cold spots in the target volume so that a more uniform dose distribution can be achieved in the target while sparing more sensitive structures.

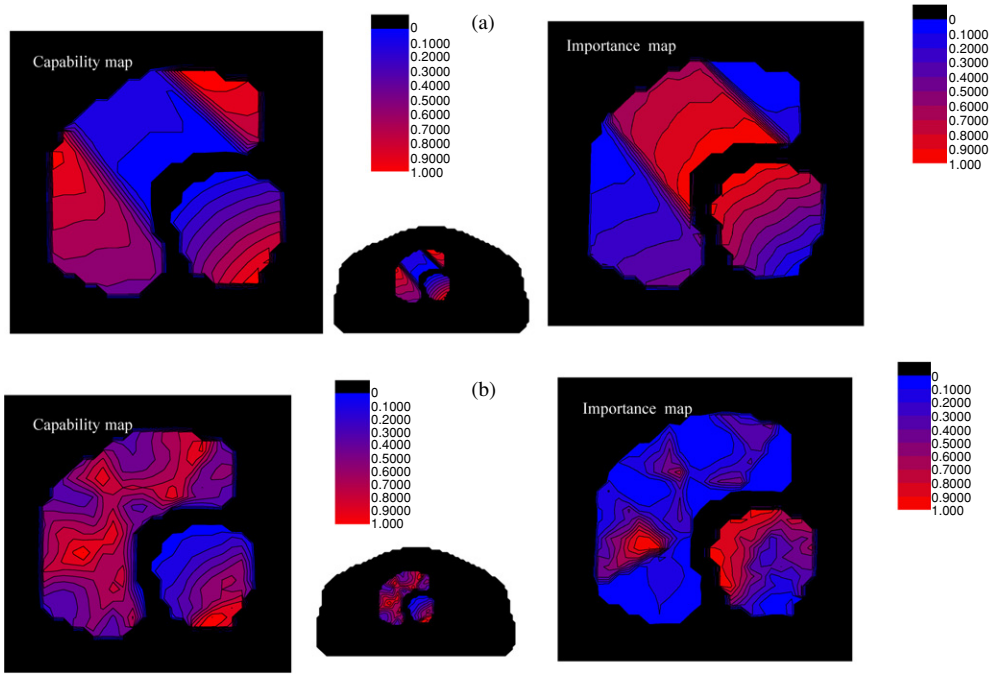


Figure 1. Dosimetric capability and importance maps of the target and sensitive structure for a hypothetical case for two beam configurations: (a) the single beam and (b) five equally spaced beams. The data for each structure are normalized to unity. For visual purposes, the capability and importance maps of the target and sensitive structure are enlarged and shown in the left and right panels for each beam configuration. The lower middle panel of each set of figures shows the complete geometry of the hypothetical structures.

2.3. Implementation

We implemented a software module to optimize the system in the platform of the PLUNC treatment planning system (an open source treatment planning system from the University of North Carolina, Chapel Hill, NC). The dose calculation engine and varieties of plan evaluation tools of the PLUNC system are used to evaluate and compare the optimization results. The SIITP (Xing and Chen 1996) was employed to obtain the optimal beam intensity profiles. The final IMRT plan is obtained in a similar manner to conventional inverse planning except that the uniform importance for the involved structures is replaced by the non-uniform importance distributions given by equations (4) and (5). The calculation was performed on a PC with P4 1.7 GHz and 1024 MB RAM.

2.4. Test of the new dose optimization formalism

To better understand the physics behind the capability calculation, we first constructed a hypothetical surrogate case (figure 1) and studied the behaviour of the system using a single beam and five incident beams. The gantry angle used for target irradiation in the single beam case was 320°, and in the five-beam case the angles used were 32°, 104°, 176°, 248° and 320°, where IEC convention for gantry angle is used. The incident photon energy was 15 MV. In both cases, the target was prescribed to 100 (arbitrary units) and the sensitive structure tolerance was set to be 25. An IMRT treatment was also planned with the same

Table 1. Summary of the parameters used for planning the IMRT prostate treatment. The tolerances of the sensitive structures are used in the evaluation of the capability maps.

	Prostate	Bladder	Rectum	Femoral heads	Skin
Importance factor	0.2	0.05	0.1	0.05	0.6
Prescription/tolerance	100	65	60	70	85

Table 2. Summary of the parameters used for planning the IMRT treatment of the paraspinal tumour. The tolerances of the involved sensitive structures are used in the evaluation of the capability maps.

	GTV	Spinal cord	Liver	Kidney	Tissue
Importance factor	0.86	0.03	0.005	0.05	0.055
Prescription/tolerance	100	30	40	30	75

five-beam configuration but structurally uniform importance factors, and the result was compared with the newly obtained plan. To ensure a fair plan comparison, in this and following examples the importance factors were chosen in such a way that the target DVHs were the same for the cases with uniform and non-uniform importance factors. The net improvements can then be assessed by the doses given to the sensitive structures.

The new algorithm was also applied to two clinical cases: a prostate case and a paraspinal tumour treatment. To illustrate the advantage of the technique, the results were compared with those obtained using the conventional inverse planning with uniform importance factors. For the prostate IMRT case, six equally spaced beams starting from 0° were used. Some relevant parameters used for planning the patient are summarized in table 1. As in conventional inverse planning, the structure-specific importance factors, $\{r_\sigma\}$, were determined by trial-and-error. To assess the new technique, we planned the case using three different strategies: (i) uniform importance for the prostate target and the sensitive structures; (ii) dosimetric capability-based non-uniform importance for the prostate target and uniform importance for all sensitive structures and (iii) dosimetric capability-based non-uniform importance for both prostate target and the sensitive structures. The three plans were compared using DVHs and isodose distribution plots.

In the paraspinal case, the sensitive structures involved include the spinal cord, liver and kidney. In this study, five 15 MV non-equally spaced coplanar beams (40°, 110°, 180°, 255° and 325°) were used for the treatment. The structure-specific importance factors and prescription/tolerances are summarized in table 2. IMRT plans with and without importance factor modulation were obtained and quantitative comparison was performed.

3. Results and discussions

3.1. A hypothetical test case

The capability and importance distributions for both single beam and five-beam configurations are plotted in figure 1. The isocapability and iso-importance curves are normalized to unity. For the single beam calculation (figure 1(a)), it is seen from the capability map (left panel) that the target region in the middle of the incident beam has lower dosimetric capability due to the restriction of the tolerance of the sensitive structure. For the points on both sides of the central target region, the dosimetric capabilities are much higher because of the absence

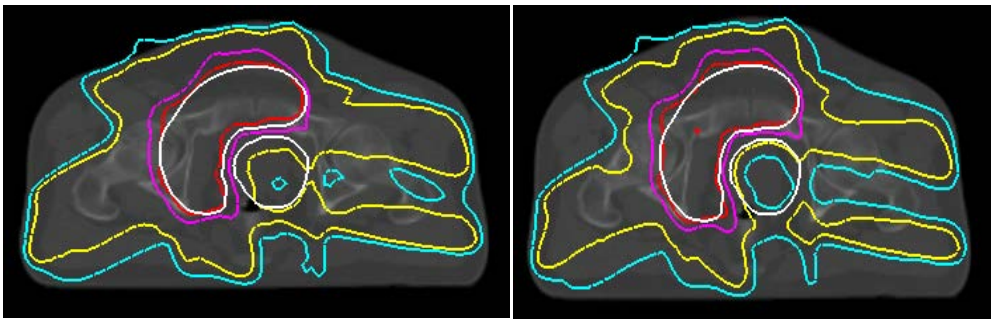


Figure 2. Isodose distributions for plans obtained without (left) and with (right) local importance factor modulation. The relative isodose curves labelled in the plots are, from the centre, 105% (red), 100% (pink), 80% (yellow) and 40% (blue), respectively.

of sensitive structure 'blocking'. The capability distribution within the sensitive structure can also be intuitively interpreted. For the voxels distant from the target volume, the capability is relatively high, indicating that these voxels are less dose-limiting points in comparison with the voxels close to the target. For the single beam case, the importance map (right panel of figure 1(a)) is almost an inversion of the capability map.

Figure 1(b) shows the capability and importance distributions when five incident beams are used to irradiate the target. The calculation is fairly efficient; it took less than 6 min for the system to obtain the capability and importance maps. In this case, the low capability region in the middle of the target shown in figure 1(a) disappears and only a few isolated low capability spots show up near the edge of the target. On the other hand, the overall behaviour of the capability map in the sensitive structure is not changed dramatically. The dosimetric capabilities for those points close to the target remain relatively low, which is consistent with our intuition since, relatively speaking, these points are more dose-limiting compared to the points far away from the target. The results also suggest that the boundary region between the two structures is likely to be underdosed (for target) or overdosed (for sensitive structure). In order to improve this, a non-uniform penalty scheme derived *a priori* or *a posteriori* becomes necessary. The technique described in this paper represents an example of the *a priori* method, whereas an adaptive adjustment of the local importance factor distribution would be an example of the latter. In general, the importance distribution is not a simple inversion of the capability map and depends also on the absolute values of the capabilities, as suggested by equations (4) and (5). This is especially true for the target, in which we need not only to 'boost' the potential cold region(s) but also to 'suppress' the potential hot spot(s). The high importance in the left-middle region of the target (see the right panel of figure 1(b)) is a direct consequence of the stated requirement. Indeed, a careful examination of the target capability data indicated that the region has high dosimetric capability and is thus likely to be overdosed. The assignment of higher local importance according to equation (4) permits us to suppress the potential overdosing *a priori*.

In figure 2, we show the IMRT plans obtained without and with modulating the spatial importance distribution. The left panel is the conventional IMRT plan with uniform importance factors assigned to both target and sensitive structures. The right panel shows the plan obtained with the spatially non-uniform importance distribution plotted in the right panel of figure 1(b). It is clearly seen that the isodose curves in the sensitive structure are 'pushed' towards the target and the dose gradient at the tumour boundary is greatly increased. The significant improvement can also be seen in the DVH plot (figure 3). It is remarkable that, by simply modulating the

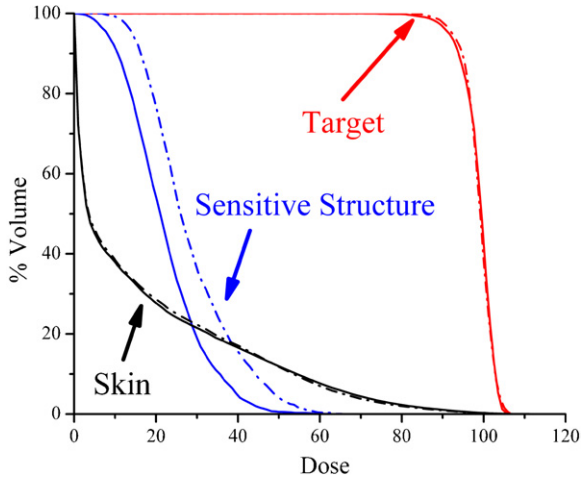


Figure 3. Target and sensitive structure DVHs corresponding to plans with two different penalty schemes: (a) uniform importance for every structure (dotted-dash curves) and (b) non-uniform importance for both target and sensitive structures (solid curves).

spatial importance distribution, an almost uniform reduction of ∼20% (normalized to the maximum sensitive structure dose) in the dose to the sensitive structure can be accomplished. If the dose to the non-sensitive structure normal tissue is not a limiting factor, the above result suggests that the target dose can be escalated by ∼10% while keeping the radiation toxicity at the current IMRT level.

3.2. Six-field IMRT prostate treatment

The IMRT plans obtained using three different penalty schemes are summarized in figures 4 and 5. The isodose distributions for the three penalty schemes are shown in figure 4. In figure 5, we compare the DVHs obtained using conventional inverse planning (dotted curves—obtained using uniform importance for the prostate target and the sensitive structures) and the new technique with non-uniform importance in both target and sensitive structures (solid curves). The DVHs of the plan with non-uniform importance to the prostate target and uniform importance to the sensitive structure are shown in dash-dotted curves. The importance factors for both target and sensitive structures were constructed based on the computed capability maps using the procedure described in section 2. It is seen from figure 4 that the dose sparing of the rectum and bladder is significantly improved when the non-uniform penalty scheme is employed. Remarkably, the maximum dose to the rectum is reduced from 68 to 60 and the fractional volume is dramatically reduced in the whole dose range. The reduction in the low dose region is more distinct, but the decrease in the high dose region is also evident and perhaps more clinically relevant. Similarly, significant improvements are achieved in the doses to bladder and femoral heads. For example, the fraction volume receiving a dose of 35 is decreased from 11% to 8% for bladder. The high dose tail in bladder is also evidently suppressed (from 70 to 66). Interestingly, the dose uniformity in the prostate target is also improved: the maximum dose of prostate target is slightly reduced from 110 to 107 and the minimum dose is increased from 85 to 87. In optimization, it is generally true that there is no net gain (that is, an improvement in the dose to a structure is often accompanied by dosimetrically adverse effect(s) at other points in the same or different structures). However,

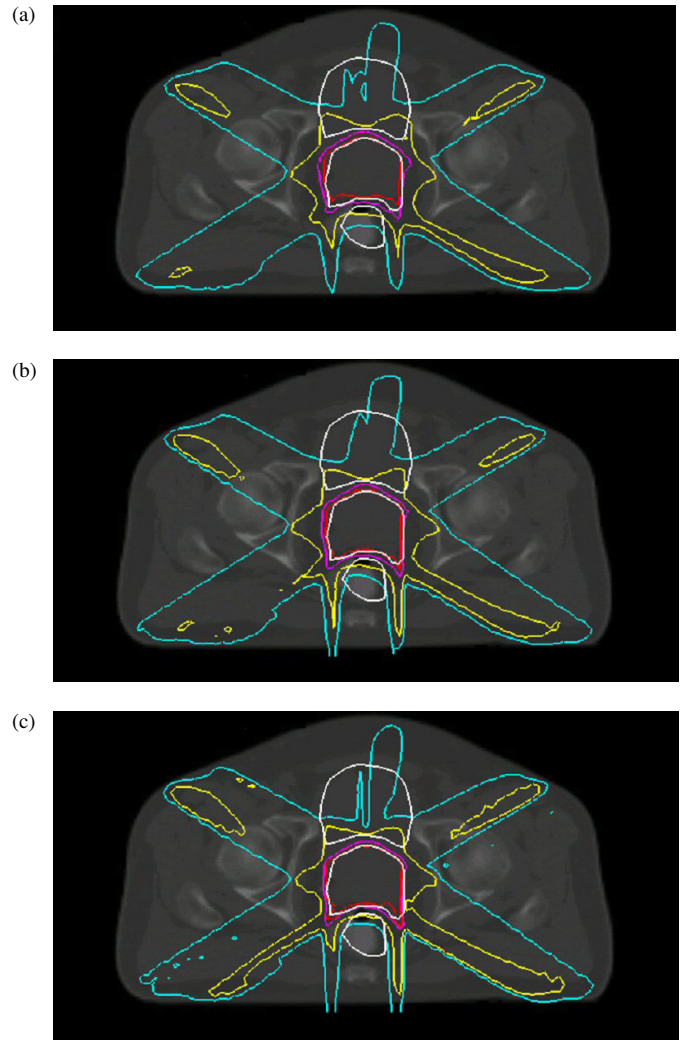


Figure 4. Isodose distributions obtained from three different penalty schemes: (a) uniform importance for every structure; (b) non-uniform importance for the prostate target and uniform importance for other structures and (c) non-uniform importance for every structure. From the centre, (red, pink, yellow and blue) curves represent 95%, 80%, 50% and 35% isodose curves. The 100% isodose curve corresponds to a dose of 72 Gy in this case.

one should note that things may be different when different penalty schemes are used, or more generally, when different objective functions are used. The simultaneous improvements in both target and sensitive structures here are a direct consequence of the enlarged solution space when non-uniform importance is permissible.

To better understand the technique, we have also optimized the dose under the condition that the non-uniform importance is allowed only for the prostate target (i.e., the importance for all sensitive structures is kept uniform). The corresponding isodose distribution is shown in figure 4(b) and the DVHs are plotted in figure 5 as dotted-dash curves. In this case, it is found that the target dose homogeneity is slightly improved. For example, the fraction receiving over 95 is increased from 91% to 96%. The maximum dose is reduced from 110 to 105.

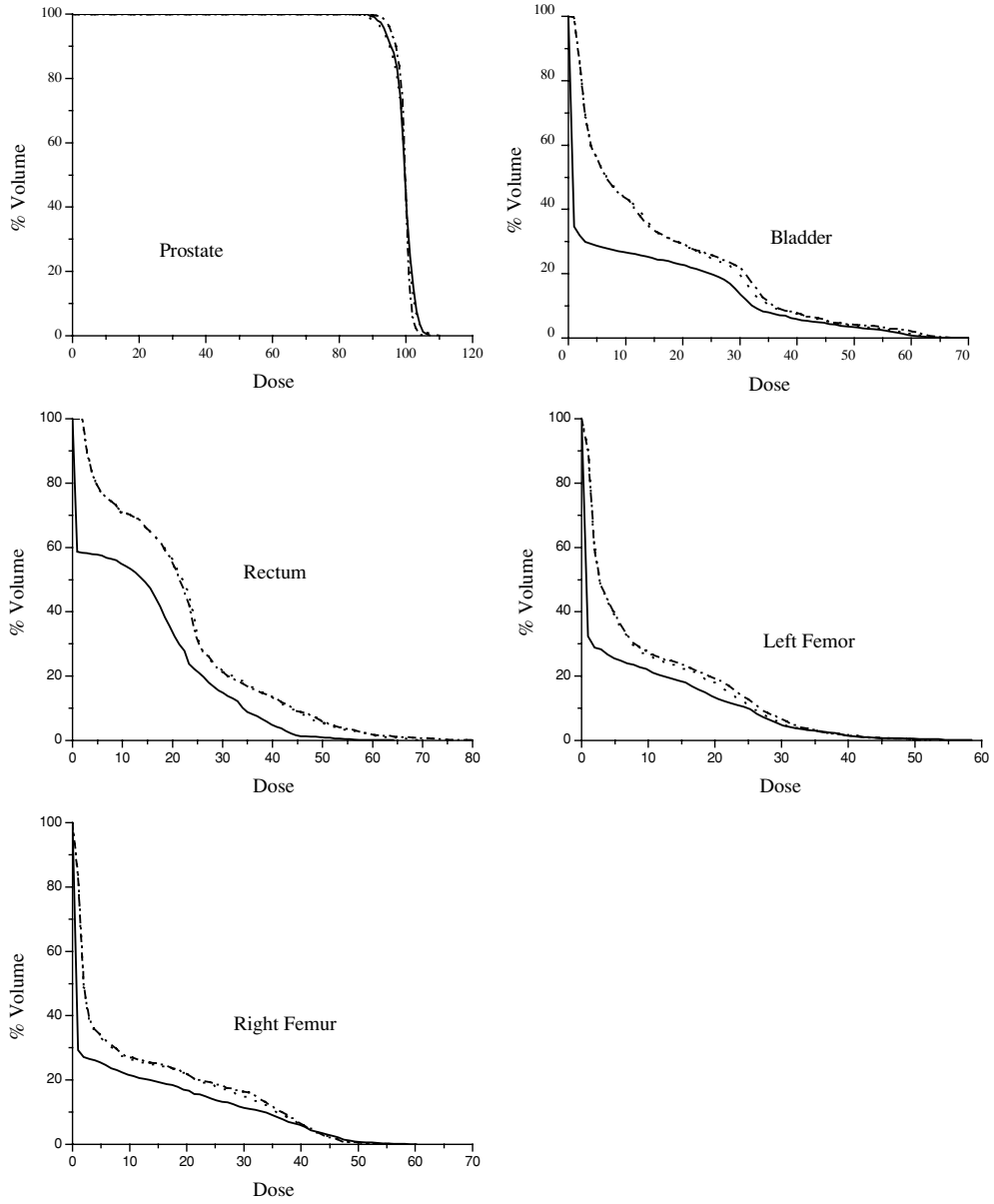


Figure 5. Target and sensitive structure DVHs corresponding to plans with three different penalty schemes: (a) uniform importance for every structure (dotted curves); (b) non-uniform importance for all structures (solid curves) and (c) non-uniform importance to the prostate and uniform importance to the sensitive structure (dash-dotted curves).

By individualizing the importance for the target voxels, we selectively increased the penalties to those voxels that are likely to be underdosed. Consequently, the target dose coverage is improved. It is interesting to note that there is essentially no change in the DVHs of the sensitive structures. The systematic improvement in the isodose distributions when the three different penalty schemes are employed can also be easily appreciated from figure 4.

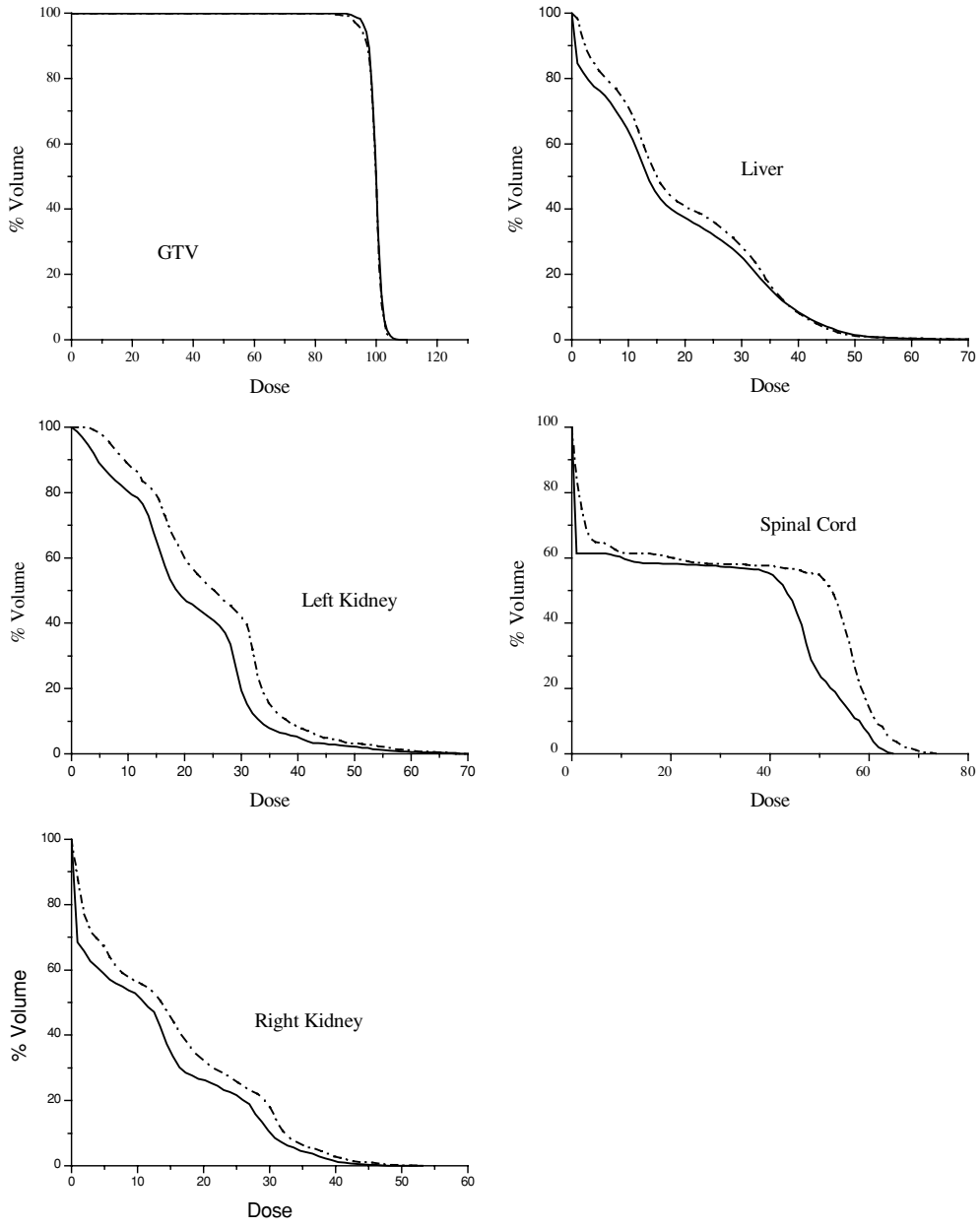


Figure 6. The comparison of DVHs for paraspinal tumour case between plan obtained from the algorithm proposed, denoted by solid line, and plan from conventional optimization, denoted by dotted dash line. The dose sparing of spinal cord, kidney and liver is evidently improved. The 100% in the x-axis corresponds to a dose of 56 Gy in this case.

3.3. Five-field IMRT treatment of a paraspinal tumour

The DVHs obtained using the conventional and newly proposed IMRT planning techniques are plotted in figure 6. The isodose distributions for the two plans are shown in figure 7. Similar

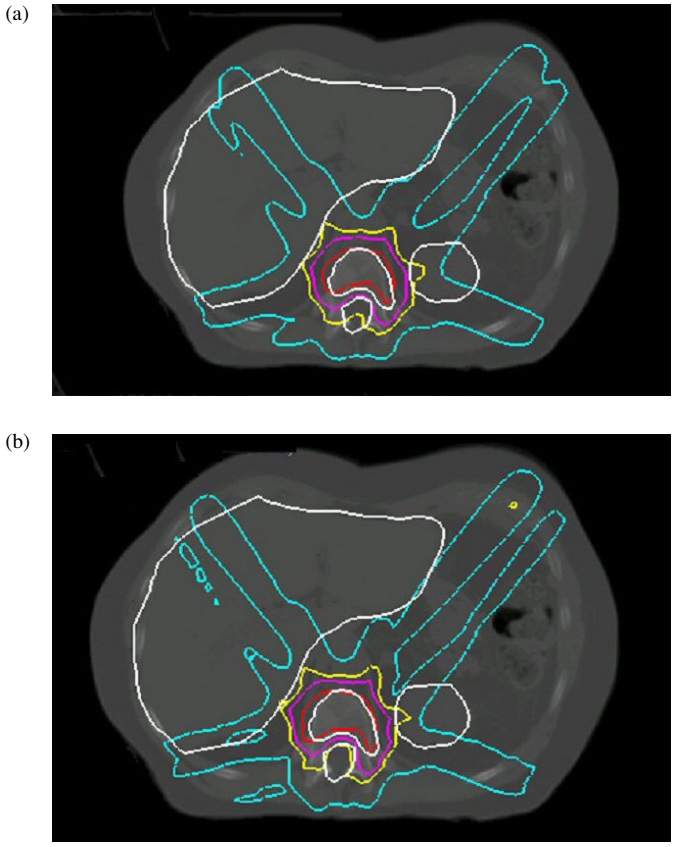


Figure 7. Isodose distributions for plans obtained with two different penalty schemes: (a) uniform importance for every structure and (b) non-uniform importance for every structure. From the centre, (red, pink, yellow and blue) curves represent 95%, 70%, 55% and 30% isodose curves.

to the previous case, when spatially non-uniform importance factors given by equation (4) are used, the target dose coverage and sensitive structure sparing are all improved in comparison with the conventional IMRT plan with uniform importance factors. For the target, the improvement is evident especially in the dose range from 90 to 95. The fractional volume receiving a dose level of 90 is slightly increased (from 99.3% to 99.7%). A more notable change is found for the fractional volume receiving doses higher than 95 (from 95.1% to 97.5%). The minimum target dose is increased from 84 to 90. The maximum target dose is, however, slightly increased from 105 to 106.5. By using the *a priori* non-uniform penalty scheme, it is found that the doses to the sensitive structures are dramatically improved. As seen from the DVHs, the spinal cord is better spared, especially in the high dose region. For example, the fractional volume receiving a dose above 50 dropped from 53% to 24%. The fractional volumes of the left kidney receiving a dose above 30 are reduced from 42% to 19% and for the right kidney, the reduction is about 7% (from 17.7% to 10.7%). The improvement for the liver is less impressive but evident. Once again, we would like to emphasize that the huge improvement in sensitive structure sparing is achieved without significantly deteriorating the dose coverage of the tumour target.

4. Conclusions

The dosimetric inequivalence of the voxels is a fundamental feature of the system and should be considered to obtain truly optimal IMRT plans. We have introduced the concept of dosimetric capability to quantify the likelihood for a voxel to meet its dosimetric goal and developed a new inverse planning formalism with the dosimetric capability-modulated voxel-dependent penalty. Using the new formalism, we can effectively 'boost' those regions where there are potential problems in meeting the dosimetric goals. With the aim of obtaining a spatially more uniform target dose distribution, in this paper, we presented a strategy to assign a higher penalty (or high local importance) to those voxels with lower capabilities. However, it is important to note that other penalty schemes can also be constructed under the guidance of the capability map to meet a different clinical requirement. The technique provides an effective mechanism to incorporate prior knowledge of the system into the dose optimization process and enables us to better model the intra-structural tradeoff. Comparison with the conventional inverse planning technique indicated that the algorithm is capable of generating much improved treatment plans with more conformal dose distributions that would otherwise be unattainable. Finally, we mention that the technique may find applications in dose optimization for many other radiation therapy modalities, such as prostate implantation, gamma knife, micro-MLC based stereotactic radiosurgery and other variants of IMRT, such as tomotherapy and intensity-modulated arc therapy.

Acknowledgments

We would like to thank Dr C Cameron for carefully reviewing the manuscript. The support of the Department of Defense (DAMD17-03-1-0657), the National Cancer Institute (5R01CA98523-02) and the Vadasz Family Foundation are gratefully acknowledged. The work of DL is supported in part by the NSF under Career Grant DMS-013351.

References

Bortfeld T, Burkelbach J, Boesecke R and Schlegel W 1990 Methods of image reconstruction from projections applied to conformation radiotherapy *Phys. Med. Biol.* **35** 1423–34

Brahme A, Roos J E and Lax I 1982 Solution of an integral equation encountered in rotation therapy *Phys. Med. Biol.* **27** 1221–9

Chen Y, Michalski D, Houser C and Galvin J M 2002 A deterministic iterative least-squares algorithm for beam weight optimization in conformal radiotherapy *Phys. Med. Biol.* **47** 1647–58

Cho P S, Lee S, Marks R J II, Oh S, Sutlief S G and Phillips M H 1998 Optimization of intensity modulated beams with volume constraints using two methods: cost function minimization and projections onto convex sets *Med. Phys.* **25** 435–43

Cotrutz C and Xing L 2002 Using voxel-dependent importance factors for interactive DVH-based dose optimization *Phys. Med. Biol.* **47** 1659–69

Cotrutz C and Xing L 2003 IMRT dose shaping using regionally variable penalty scheme *Med. Phys.* **30** 544–51

Das S, Cullip T, Tracton G, Chang S, Marks L, Anscher M, Rosenman J and Chang S X 2003 Beam orientation selection for intensity-modulated radiation therapy based on target equivalent uniform dose maximization *Int. J. Radiat. Oncol. Biol. Phys.* **55** 215–24

Gopal R and Starkschall G 2001 Plan space: representation of treatment plans in multidimensional space *Med. Phys.* **28** 1227

Holmes T and Mackie T R 1994 A comparison of three inverse treatment planning algorithms *Phys. Med. Biol.* **39** 91–106

Hristov D H and Fallone B G 1997 An active set algorithm for treatment planning optimization *Med. Phys.* **24** 1455–64

Langer M, Brown R, Urie M, Leong J, Stracher M and Shapiro J 1990 Large scale optimization of beam weights under dose–volume restrictions *Int. J. Radiat. Oncol. Biol. Phys.* **18** 887–93

Langer M and Leong J 1987 Optimization of beam weights under dose–volume restrictions *Int. J. Radiat. Oncol. Biol. Phys.* **13** 1255–60

Lian J and Xing L 2004 Incorporating model parameter uncertainty into inverse treatment planning *Med. Phys.* **31** 2711–20

Michalski D, Xiao Y, Censor Y and Galvin J 2004 The dose–volume constraint satisfaction problem for inverse treatment planning with field segments *Phys. Med. Biol.* **49** 601–16

Oelfke U and Bortfeld T 1999 Inverse planning for x-ray rotation therapy: a general solution of the inverse problem *Phys. Med. Biol.* **44** 1089–104

Olivera G H, Shepard D M, Reckwerdt P J, Ruchala K, Zachman J, Fitchard E E and Mackie T R 1998 Maximum likelihood as a common computational framework in tomotherapy *Phys. Med. Biol.* **43** 3277–94

Rosen I I, Lam K S, Lane R G, Langer M and Morrill S M 1995 Comparison of simulated annealing algorithms for conformal therapy treatment planning *Int. J. Radiat. Oncol. Biol. Phys.* **33** 1091–9

Spirou S V and Chui C S 1998 A gradient inverse planning algorithm with dose–volume constraints *Med. Phys.* **25** 321–33

Stark H and Yang Y 1988 *Vector Space Projections: A Numerical Approach to Signal and Image Processing, Neural Nets, and Optics* (New York: Wiley)

Starkschall G and Eifel P J 1992 An interactive beam-weight optimization tool for three-dimensional radiotherapy treatment planning *Med. Phys.* **19** 155–63

Thieke C, Bortfeld T, Niemierko A and Nill S 2003 From physical dose constraints to equivalent uniform dose constraints in inverse radiotherapy planning *Med. Phys.* **30** 2332–9

Webb S 1991 Optimization by simulated annealing of three-dimensional conformal treatment planning for radiation fields defined by a multileaf collimator *Phys. Med. Biol.* **36** 1201–26

Webb S, Convery D J and Evans P M 1998 Inverse planning with constraints to generate smoothed intensity-modulated beams *Phys. Med. Biol.* **43** 2785–94

Wu C, Olivera G H, Jeraj R, Keller H and Mackie T R 2003 Treatment plan modification using voxel-based weighting factors/dose prescription *Phys. Med. Biol.* **48** 2479–91

Wu Q, Mohan R, Niemierko A and Schmidt -Ullrich R 2002 Optimization of intensity-modulated radiotherapy plans based on the equivalent uniform dose *Int. J. Radiat. Oncol. Biol. Phys.* **52** 224–35

Wu X and Zhu Y 2001 An optimization method for importance factors and beam weights based on genetic algorithms for radiotherapy treatment planning *Phys. Med. Biol.* **46** 1085–99

Xiao Y, Galvin J, Hossain M and Valicenti R 2000 An optimized forward-planning technique for intensity modulated radiation therapy *Med. Phys.* **27** 2093–9

Xing L and Chen G T Y 1996 Iterative algorithms for inverse treatment planning *Phys. Med. Biol.* **41** 2107–23

Xing L, Li J G, Donaldson S, Le Q T and Boyer A L 1999 Optimization of importance factors in inverse planning *Phys. Med. Biol.* **44** 2525–36

Zagars G K, Starkschall G, Antolak J A, Lee J J, Huang E, von Eschenbach A C, Kuban D A and Rosen I 2002 Treatment planning using a dose–volume feasibility search algorithm *Int. J. Radiat. Oncol. Biol. Phys.* **53** 1097–105